



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number 121207**

**TO: Manjunath N Rao**  
**Location: rem/3b813c70**  
**Art Unit: 1652**  
**Thursday, May 13, 2004**

**Case Serial Number: 09/930440**

**From: Toby Port**  
**Location: Biotech-Chem Library**  
**Remsen 1A59**  
**Phone: 571-272-2523**

**toby.port@uspto.gov**

### **Search Notes**

Dear Examiner Rao,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor  
571-272-2507 Remsen E01 D86

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.





Db 61 ATCATTCGCCAGATCGCCAGAACACACAGGCGACCTGAGCTAGACCAAGCGCATGATC 120  
 Qy 121 CGCATGCGCCAGAGAGTGTGGGCTGATTTGTGCCAACTTCCAGAAAGAGTAAATTC 180  
 Db 121 CGCATGCGCCAGAGAGTGTGGGCTGATTTGTGCCAACTTCCAGAAAGAGTAAATTC 180  
 Qy 181 AAGTTTATCGAAAGCCTTGGAGAGGCCATACACTCGAACATTCCTGGGGAGAGC 240  
 Db 181 AAGTTTATCGAAAGCCTTGGAGAGGCCATACACTCGAACATTCCTGGGGAGAGC 240  
 Qy 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCATGACCACTAGAGGAGCTGAGAG 300  
 Db 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCATGACCACTAGAGGAGCTGAGAG 300  
 Qy 301 TACGGCGAGAGGATGGATCTTCTTCACTGCTCTGAGATGAGATGAGAGGATGAA 360  
 Db 301 TACGGCGAGAGGATGGATCTTCTTCACTGCTCTGAGATGAGATGAGAGGATGAA 360  
 Qy 361 TTCTTGATGAACTGAAATGTTTCCATTTTCAAGTTGAACTGAGAGACCTAATATTTT 420  
 Db 361 TTCTTGATGAACTGAAATGTTTCCATTTTCAAGTTGAACTGAGAGACCTAATATTTT 420  
 Qy 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATCTCCAGTGGAGTGGAG 480  
 Db 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATCTCCAGTGGAGTGGAG 480  
 Qy 481 TCATGAGACACCAATGAGAGAGTTTATCAGATGTTGAGAGCCCTCAACCCCACTTCTGC 540  
 Db 481 TCATGAGACACCAATGAGAGAGTTTATCAGATGTTGAGAGCCCTCAACCCCACTTCTGC 540  
 Qy 541 TTCTTGCAATGTAACGAGCGCATACCGCTCCAGCCCTGAGAGAGTCAACCTGGGGTCAATC 600  
 Db 541 TTCTTGCAATGTAACGAGCGCATACCGCTCCAGCCCTGAGAGAGTCAACCTGGGGTCAATC 600  
 Qy 601 TCGGAATATCAGAAAGCTTCTTCCGATCTTCCCATATGAGGATCTTCTGGCATGAAACAGGC 660  
 Db 601 TCGGAATATCAGAAAGCTTCTTCCGATCTTCCCATATGAGGATCTTCTGGCATGAAACAGGC 660  
 Qy 661 ATAGCGATATCTGTGGCCGAGTGGTCTTGGGGGCGCAAGGTGTGGAAGTCACTACT 720  
 Db 661 ATAGCGATATCTGTGGCCGAGTGGTCTTGGGGGCGCAAGGTGTGGAAGTCACTACT 720  
 Qy 721 TTGGAACAAGCTGGAAGGAGAGTGAACAACCTGAGCTGAGAGCTGAGAGCTGAGC 780  
 Db 721 TTGGAACAAGCTGGAAGGAGAGTGAACAACCTGAGCTGAGAGCTGAGAGCTGAGC 780  
 Qy 781 GAGCTGAGTGGGTCAGTGGCTCTTGTGAGAGCTGGCTCCCAACCAAGAGCTG 840  
 Db 781 GAGCTGAGTGGGTCAGTGGCTCTTGTGAGAGCTGGCTCCCAACCAAGAGCTG 840  
 Qy 841 CTGCGCTGAGATGAGGCTGCAATGAGAGTGGGGAAGTCTGTGCTGAGCAAGTGA 900  
 Db 841 CTGCGCTGAGATGAGGCTGCAATGAGAGTGGGGAAGTCTGTGCTGAGCAAGTGA 900  
 Qy 901 ATTCCGGAAGGACCAATTTCAATGAGACATGCTCAACCGTGAAGGTGGTGAAGCCAAA 960  
 Db 901 ATTCCGGAAGGACCAATTTCAATGAGACATGCTCAACCGTGAAGGTGGTGAAGCCAAA 960  
 Qy 961 GCTTATCTGCTGAGAGACATTTTATCTGAGTGGGAGAGAGTCTGTGCTGAGTGA 1020  
 Db 961 GCTTATCTGCTGAGAGACATTTTATCTGAGTGGGAGAGAGTCTGTGCTGAGTGA 1020  
 Qy 1021 GAGGATGACACCATCTAGAGAGATTTGATGATTAATCATGCAAAAAATCAAGTCTTA 1080  
 Db 1021 GAGGATGACACCATCTAGAGAGATTTGATGATTAATCATGCAAAAAATCAAGTCTTA 1080

RESULT 2  
 US-09-930-440B-5  
 Sequence 5, Application US/09930440B  
 Patent No. US20020142386A1  
 GENERAL INFORMATION:  
 APPLICANT: Betendbaugh et al.

TITLE OF INVENTION: Engineering Intracellular Signaling Pathways  
 FILE REFERENCE: PFO92  
 CURRENT APPLICATION NUMBER: US/09/930,440B  
 CURRENT FILING DATE: 2001-08-16  
 PRIOR APPLICATION NUMBER: 60/227,579  
 PRIOR FILING DATE: 2000-08-25  
 PRIOR APPLICATION NUMBER: 09/516,793  
 PRIOR FILING DATE: 2000-03-01  
 PRIOR APPLICATION NUMBER: 60/169,624  
 PRIOR FILING DATE: 1999-12-08  
 PRIOR APPLICATION NUMBER: 60/122,582  
 PRIOR FILING DATE: 1999-03-02  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 5  
 LENGTH: 1080  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1) ..(1080)  
 US-09-930-440B-5

Query Match 100.0%; Score 1080; DB 9; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCTGAGAGTGTGGGCTGATTTCCGAGCGCTGGGTGGGCGGCAACACCGTCTTC 60  
 Db 1 ATGCGCTGAGAGTGTGGGCTGATTTCCGAGCGCTGGGTGGGCGGCAACACCGTCTTC 60  
 Qy 61 ATCATTCGCCAGATCGCCAGAACACACAGGCGACCTGAGCTAGACCAAGCGCATGATC 120  
 Db 61 ATCATTCGCCAGATCGCCAGAACACACAGGCGACCTGAGCTAGACCAAGCGCATGATC 120  
 Qy 121 CGCATGCGCCAGAGAGTGTGGGCTGATTTGTGCCAACTTCCAGAAAGAGTAAATTC 180  
 Db 121 CGCATGCGCCAGAGAGTGTGGGCTGATTTGTGCCAACTTCCAGAAAGAGTAAATTC 180  
 Qy 181 AAGTTTATCGAAAGCCTTGGAGAGGCCATACACTCGAACATTCCTGGGGAGAGC 240  
 Db 181 AAGTTTATCGAAAGCCTTGGAGAGGCCATACACTCGAACATTCCTGGGGAGAGC 240  
 Qy 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCATGACCACTAGAGGAGCTGAGAG 300  
 Db 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCATGACCACTAGAGGAGCTGAGAG 300  
 Qy 301 TACGGCGAGAGGATGGATCTTCTTCACTGCTCTGAGATGAGATGAGAGGATGAA 360  
 Db 301 TACGGCGAGAGGATGGATCTTCTTCACTGCTCTGAGATGAGATGAGAGGATGAA 360  
 Qy 361 TTCTTGATGAACTGAAATGTTTCCATTTTCAAGTTGAACTGAGAGACCTAATATTTT 420  
 Db 361 TTCTTGATGAACTGAAATGTTTCCATTTTCAAGTTGAACTGAGAGACCTAATATTTT 420  
 Qy 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATCTCCAGTGGAGTGGAG 480  
 Db 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATCTCCAGTGGAGTGGAG 480  
 Qy 481 TCATGAGACACCAATGAGAGAGTTTATCAGATGTTGAGAGCCCTCAACCCCACTTCTGC 540  
 Db 481 TCATGAGACACCAATGAGAGAGTTTATCAGATGTTGAGAGCCCTCAACCCCACTTCTGC 540  
 Qy 541 TTCTTGCAATGTAACGAGCGCATACCGCTCCAGCCCTGAGAGAGTCAACCTGGGGTCAATC 600  
 Db 541 TTCTTGCAATGTAACGAGCGCATACCGCTCCAGCCCTGAGAGAGTCAACCTGGGGTCAATC 600  
 Qy 601 TCGGAATATCAGAAAGCTTCTTCCGATCTTCCCATATGAGGATCTTCTGGCATGAAACAGGC 660  
 Db 601 TCGGAATATCAGAAAGCTTCTTCCGATCTTCCCATATGAGGATCTTCTGGCATGAAACAGGC 660  
 Qy 661 ATAGCGATATCTGTGGCCGAGTGGTCTTGGGGGCGCAAGGTGTGGAAGTCACTACT 720



Thu May 13 11:53:07 2004

us-09-930-440b-5.rnpb

Page 3

Db 661 ATAGCATATCTGTGCGCCGAGTGGCTCTG9393CCAAAGGTGTGGAACTTCACATAACT 720  
Qy 721 TTGGACAAGACCTGGAAAGGGGAGTGAACCACTCGGCGCTCGTGAAGCTTGGAACTGGCC 780  
Db 721 TTGGACAAGACCTGGAAAGGGGAGTGAACCACTCGGCGCTCGTGAAGCTTGGAACTGGCC 780  
Qy 781 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGGAGCGTGCCTCGGCTTCCCAACCAAGCACTG 840  
Db 781 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGGAGCGTGCCTCGGCTTCCCAACCAAGCACTG 840  
Qy 841 CTGCGCTGTGAGATGGCTTGCATGAGAACTGGGCAAGTCTGTGTGGGCAAGTGA 900  
Db 841 CTGCGCTGTGAGATGGCTTGCATGAGAACTGGGCAAGTCTGTGTGGGCAAGTGA 900  
Qy 901 ATTCCGGAAGGACCACTTCTTAACAAATGACATGCTCAACGCTGAAGTGGTGAAGCCCAA 960  
Db 901 ATTCCGGAAGGACCACTTCTTAACAAATGACATGCTCAACGCTGAAGTGGTGAAGCCCAA 960  
Qy 961 GCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGACTGTGA 1020  
Db 961 GCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGACTGTGA 1020  
Qy 1021 GAGGATGACACCATCATTGAGAAAGATTGTGATATGATGAGCAAAATCAAGTCTTA 1080  
Db 1021 GAGGATGACACCATCATTGAGAAAGATTGTGATATGATGAGCAAAATCAAGTCTTA 1080

RESULT 3  
US-10-264-237-1192  
Sequence 1192, Application US/10264237  
Publication No. US20040005491A1  
GENERAL INFORMATION:  
APPLICANT: Bize et al.  
FILE REFERENCE: PA131P1  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
CURRENT APPLICATION NUMBER: US/10/264,237  
CURRENT FILING DATE: 2002-10-04  
PRIORITY APPLICATION NUMBER: PCT/US01/16450  
PRIORITY FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: US 60/205,515  
PRIOR FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 2876  
SOFTWARE: PatentIn Ver. 3.1  
SEQ ID NO 1192  
LENGTH: 1268  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (5)..(5)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (17)..(17)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (23)..(23)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (31)..(31)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1224)..(1227)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (1230)..(1230)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1257)..(1257)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1265)..(1265)  
OTHER INFORMATION: n equals a,t,g, or c  
US-10-264-237-1192  
Query Match 99.8%; Score 1078; DB 16; Length 1268;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1078; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCGCTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGCGGCAACCCGTCCTTC 60  
Db 121 ATGCCGCTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGCGGCAACCCGTCCTTC 180  
Qy 61 ATCAATGCGGAGATCGGCCAAGAACCAACGAGGAGACTTGAAGCTTACCAAGCTATGTC 120  
Db 181 ATCAATGCGGAGATCGGCCAAGAACCAACGAGGAGACTTGAAGCTTACCAAGCTATGTC 240  
Qy 121 CGCATGCCCAAGAGAGTGGGGCTATGTGGCAAGTTCAGAAAGATGAGTAAATTC 180  
Db 241 CGCATGCCCAAGAGAGTGGGGCTATGTGGCAAGTTCAGAAAGATGAGTAAATTC 360  
Qy 181 AAGTTTATCGAAGAACCTTGGAGAGGCCATACCTCGAAGCTTCTGGGGGAAGAG 240  
Db 301 AAGTTTATCGAAGAACCTTGGAGAGGCCATACCTCGAAGCTTCTGGGGGAAGAG 360  
Qy 241 TAGGGGAGCAACAAAGACATCTGAGTCAAGCAATGACATGACAGAGGAGTGCAGAG 300  
Db 361 TAGGGGAGCAACAAAGACATCTGAGTCAAGCAATGACATGACAGAGGAGTGCAGAG 420  
Qy 301 TAGCGGAGAGGTGGGATCTTCTTCACTGCTCTGGCATGATGATGAGATGGAGTTGAA 360  
Db 421 TAGCGGAGAGGTGGGATCTTCTTCACTGCTCTGGCATGATGATGAGATGGAGTTGAA 480  
Qy 421 CTCTATCTGAAAAAGCAGCAAAAAAGGTGCGCCCAATGGTGAATCTCGATGGGATGAG 480  
Db 541 CTCTATCTGAAAAAGCAGCAAAAAAGGTGCGCCCAATGGTGAATCTCGATGGGATGAG 600  
Qy 541 TCAATGAGACCATGAGCAAGTTTATCAGATCTGTAAGCCCTTCAACCCCACTTCTGC 540  
Db 601 TCAATGAGACCATGAGCAAGTTTATCAGATCTGTAAGCCCTTCAACCCCACTTCTGC 660  
Qy 541 TTCTTGACGTGTAACGAGCATACCGCTCAAGCTGAGAGAGTCAACCTGCGGGTCATC 600  
Db 661 TTCTTGACGTGTAACGAGCATACCGCTCAAGCTGAGAGAGTCAACCTGCGGGTCATC 720  
Qy 601 TCGGAATATCAGAAAGCTCTTCTGACATTCCTCAATGAGGTATTTCTGGGCAATGAACA 660  
Db 721 TCGGAATATCAGAAAGCTCTTCTGACATTCCTCAATGAGGTATTTCTGGGCAATGAACA 780  
Qy 661 ATAGGATATCTGTGGCGCGAGTGGCTGAGGGGCAAGGTGTGGAAGCTCATATACT 720  
Db 781 ATAGGATATCTGTGGCGCGAGTGGCTGAGGGGCAAGGTGTGGAAGCTCATATACT 840  
Qy 721 TTGGACAAGACCTGGAAGGGGAGTGAACCTCGGCTGCTGAGAGCTTGAAGAACTGGCC 780  
Db 841 TTGGACAAGACCTGGAAGGGGAGTGAACCTCGGCTGCTGAGAGCTTGAAGAACTGGCC 900  
Qy 781 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGGAGCGTGCCTCGGCTTCCCAACCAAGCACTG 840  
Db 901 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGGAGCGTGCCTCGGCTTCCCAACCAAGCACTG 960  
Qy 841 CTGCGCTGTGAGATGGCTTGCATGAGAACTGGGCAAGTCTGTGTGGGCAAGTGA 900

Db 961 CTGCGCTGTGAGTGGCTGCAATGAGAAAGCTGGCAAGCTGTGTGTGGCCAAATGAAA 1020  
Qy 901 ATTCGGGAAGGCAACATTTCTMAAATGACATGCTCAACCGTGAAGGTGGTGAGCCCAA 960  
Db 1021 ATTCGGGAAGGCAACATTTCTMAAATGACATGCTCAACCGTGAAGGTGGTGAGCCCAA 1080  
Qy 961 GCTATCTCTCTGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGACTGTGAA 1020  
Db 1081 GCTATCTCTCTGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGACTGTGAA 1140  
Qy 1021 GAGGATGACACCATCATGAGAAATGTGTAGATAATCATGCAAAAAATCAAGTCTTAA 1080  
Db 1141 GAGGATGACACCATCATGAGAAATGTGTAGATAATCATGCAAAAAATCAAGTCTTAA 1200

RESULT 4  
US-10-037-270-1035  
Sequence 1035, Application US/10037270  
Publication No. US20030104529A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyun  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Weinman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yundong  
APPLICANT: Wang, Dunhui  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillinghast, John  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1 Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_Fl\_genes Version 1.0  
SEQ ID NO 1035  
LENGTH: 1230  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (119) ..(1198)  
US-10-037-270-1035

Query Match 99.7%; Score 1076.8; DB 15; Length 1230;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCGTGTGAGTGGCTGCAATGAGAAAGCTGGCAAGCTGTGTGTGGCCAAATGAAA 60  
Db 119 ATGCGCGTGTGAGTGGCTGCAATGAGAAAGCTGGCAAGCTGTGTGTGGCCAAATGAAA 178  
Qy 61 ATCATTGCGGAGATCGGCGCAACCAACCGGCGACCTGGAGTGAAGCGCATGATC 120  
Db 179 ATCATTGCGGAGATCGGCGCAACCAACCGGCGACCTGGAGTGAAGCGCATGATC 238  
Qy 121 CGCATGCGCAAGAGAGTGGGCTGATTTGTGCAAGTTCCAGAAAGTGTAGATTC 180  
Db 239 CGCATGCGCAAGAGAGTGGGCTGATTTGTGCAAGTTCCAGAAAGTGTAGATTC 298

Qy 181 AAGTTATCGAAAGGCTTGGAGAGGCAATACACCTGTGAAAGTCTCTGTGGGAAAGAG 240  
Db 299 AAGTTATCGAAAGGCTTGGAGAGGCAATACACCTGTGAAAGTCTCTGTGGGAAAGAG 358  
Qy 241 TACCGGAGGCAAAAGCAATCTGGAATTCAGCCATTAACCAATGACAGGAGCTGCAGAG 300  
Db 359 TACCGGAGGCAAAAGCAATCTGGAATTCAGCCATTAACCAATGACAGGAGCTGCAGAG 418  
Qy 301 TACCGGAGGAGTGGGATCTTCTTCACTGCTCTGAGTGAAGTGAAGTGAAGTGAAG 360  
Db 419 TACCGGAGGAGTGGGATCTTCTTCACTGCTCTGAGTGAAGTGAAGTGAAGTGAAG 478  
Qy 361 TTCCGTGATGAGTGAATGTTTCAATTTTCAAGTTGATGAGTGAAGTGAAGTGAAG 420  
Db 479 TTCCGTGATGAGTGAATGTTTCAATTTTCAAGTTGATGAGTGAAGTGAAGTGAAG 538  
Qy 421 CCTATCTGGAAGGCAAGGCAAAAGGTCGCCCAATGATCTCAGTGGAGTGCAG 480  
Db 539 CCTATCTGGAAGGCAAGGCAAAAGGTCGCCCAATGATCTCAGTGGAGTGCAG 598  
Qy 481 TCAATGACACCATGAAAGAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540  
Db 599 TCAATGACACCATGAAAGAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 658  
Qy 541 TTCTTGCAAGTGTACCAAGGCAATCCGCTCAAGCTGAGAGCTGAGGAGTGCATC 600  
Db 659 TTCTTGCAAGTGTACCAAGGCAATCCGCTCAAGCTGAGAGCTGAGGAGTGCATC 718  
Qy 601 TCGGAATATCAGAAAGCTTTTCTGCAATTCCTATAGGATATCTGGGCAAGAAAGGC 660  
Db 719 TCGGAATATCAGAAAGCTTTTCTGCAATTCCTATAGGATATCTGGGCAAGAAAGGC 778  
Qy 661 ATAGCATATCTGTGGCGGCAAGTGTCTGTGGGCAAGTGTGTGAAAGTGTGAAAGT 720  
Db 779 ATAGCATATCTGTGGCGGCAAGTGTCTGTGGGCAAGTGTGTGAAAGTGTGAAAGT 838  
Qy 721 TTGGAAGAAGCTTGAAGGAGTGAACCACTGGGCTGTGAGGCTGGAAGTGGCC 780  
Db 839 TTGGAAGAAGCTTGAAGGAGTGAACCACTGGGCTGTGAGGCTGGAAGTGGCC 898  
Qy 781 GAGTGTGTGAGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840  
Db 899 GAGTGTGTGAGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 958  
Qy 841 CTGCGCTGTGAGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900  
Db 959 CTGCGCTGTGAGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1018  
Qy 901 ATTCGGGAAGGCAACATTTCTMAAATGACATGCTCAACCGTGAAGGTGGTGAGCCCAA 960  
Db 1019 ATTCGGGAAGGCAACATTTCTMAAATGACATGCTCAACCGTGAAGGTGGTGAGCCCAA 1078  
Qy 961 GCTATCTCTCTGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGACTGTGAA 1020  
Db 1079 GCTATCTCTCTGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGACTGTGAA 1138  
Qy 1021 GAGGATGACACCATCATGAGAAATGTGTAGATAATCATGCAAAAAATCAAGTCTTAA 1080  
Db 1139 GAGGATGACACCATCATGAGAAATGTGTAGATAATCATGCAAAAAATCAAGTCTTAA 1198

RESULT 5  
US-10-117-722-1035  
Sequence 1035, Application US/10117722  
Publication No. US20030219744A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030219744A1 Nucleic Acids and

```

1  TITLE OF INVENTION: Polypeptides
2  FILE REFERENCE: 784CIP28CIP
3  CURRENT APPLICATION NUMBER: US/0/117,722
4  CURRENT FILING DATE: 2002-04-04
5  PRIOR APPLICATION NUMBER: 09/620,312
6  PRIOR FILING DATE: 2000-07-19
7  PRIOR APPLICATION NUMBER: 09/552,317
8  PRIOR FILING DATE: 2000-04-25
9  PRIOR APPLICATION NUMBER: 09/488,725
10 PRIOR FILING DATE: 2000-01-21
11 NUMBER OF SEQ ID NOS: 1104
12 SOFTWARE: pt FL_genes Version 1.0
13 SEQ ID NO 1035
14 LENGTH: 1230
15 TYPE: DNA
16 ORGANISM: Homo sapiens
17 FEATURE:
18 NAME/KEY: CDS
19 LOCATION: (119)..(1198)
20 US-10-117-722-1035

```

Query Match	99.7%;	Score 1076.8;	DB 16;	Length 1250;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1078; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	1	ATGCGCGCTGGAGCTGGAGCTGTATCCCGGGCGCTGGGTGGCGGGCAACACCCGGCTTC	60
Db	119	ATGCGCGCTGGAGCTGGAGCTGTATCCCGGGCGCTGGGTGGCGGGCAACACCCGGCTTC	178
QY	61	ATCATTTGCCAGATCGCGCAAGACCACAGGGCGACTGGACGTAGCCAGCGCATATC	120
Db	179	ATCATTTGCCAGATCGCGCAAGACCACAGGGCGACTGGACGTAGCCAGCGCATATC	238
QY	121	CGCATGGCCAGAGAGTGTGGGGCTGATTTGTGGCAAGTTCCAGAAAGATGAGCTTAATTC	180
Db	239	CGCATGGCCAGAGAGTGTGGGGCTGATTTGTGGCAAGTTCCAGAAAGATGAGCTTAATTC	298
QY	181	AAGTTAATTCGGAAGCCTTTGGAGAGCCATACACTTCGAGACATTCCTGGGGGAAGAG	240
Db	299	AAGTTAATTCGGAAGCCTTTGGAGAGCCATACACTTCGAGACATTCCTGGGGGAAGAG	358
QY	241	TACGGGGAGCAACAAACGACATCTGGAGTTCAAGCCATGACGATACAGAGAGAGTGGAGAG	300
Db	359	TACGGGGAGCAACAAACGACATCTGGAGTTCAAGCCATGACGATACAGAGAGAGTGGAGAG	418
QY	301	TACGCCAGAGAGTGGAGATCTTCTTCACTGCTCTGGCATAGATGATGGCACTTGA	360
Db	419	TACGCCAGAGAGTGGAGATCTTCTTCACTGCTCTGGCATAGATGATGGCACTTGA	478
QY	361	TTCTCGCATGAAGTGAATGTTCCATTTTCAAGTTGGATCTGGAGACACTATATATTT	420
Db	479	TTCTCGCATGAAGTGAATGTTCCATTTTCAAGTTGGATCTGGAGACACTATATATTT	538
QY	421	CCTTATCTGGAAAAAGACAGCCAAAAAAGTGGCCCATAGGTGATCTCCAGTGGAGTGAAG	480
Db	539	CCTTATCTGGAAAAAGACAGCCAAAAAAGTGGCCCATAGGTGATCTCCAGTGGAGTGAAG	598
QY	481	TCAATGGACACATGAGCAAGTTTATCGATGTGAAGCCCTCAACCCCAACTTGTGC	540
Db	599	TCAATGGACACATGAGCAAGTTTATCGATGTGAAGCCCTCAACCCCAACTTGTGC	658
QY	541	TTCTTGGAGTACACAGGCGATACCCGCTCCAGCCTGGAGAGGTCAACCTGGGGTCAATC	600
Db	659	TTCTTGGAGTACACAGGCGATACCCGCTCCAGCCTGGAGAGGTCAACCTGGGGTCAATC	718
QY	601	TCGGATAATCAGAAAGCTTTTCTTGACATTCCTCATAGGATATTTCTGGCATGAAAACAGGC	660
Db	719	TCGGATAATCAGAAAGCTTTTCTTGACATTCCTCATAGGATATTTCTGGCATGAAAACAGGC	778
QY	661	ATAAGCGAATATCTGGCGCGGAGTGGCTCTGGGGCGCAAGGTGTGGAACGTCACTAATCT	720
Db	779	ATAAGCGAATATCTGGCGCGGAGTGGCTCTGGGGCGCAAGGTGTGGAACGTCACTAATCT	838

[illegible]

```

RESULT 6
US-09-880-107-1514/c
Sequence 1514, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1514
LENGTH: 464
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA621146
US-09-880-107-1514

```

	Query Match	38.7%	Score 418;	DB 9;	Length 464;
	Best Local Similarity	98.6%	Pred. No. 4	4e-124;	
	Matches 432;	Conservative	0;	Mismatches 5;	Indels 1;
				Gaps	1;
QY	643	TCCTGGGCATGAAACAGGCATAGCCGATATCTGTGGCCGACAGTGGCTCTTGCGGGCCCAAGTGT	702		
Db	464	TCCTGGGCATGAAACAGGCATAGCCGATATCTGTGGCCGACAGTGGCTCTTGCGGGCCCAAGTGT	405		
QY	703	TTGGAAAGCTACATTAATCTTGGACAAGACTGGAAAGGAGAGTGAACATCGAGCTCCGCTGT	762		
Db	404	TTGGAAAGCTACATTAATCTTGGACAAGACTGGAAAGGAGAGTGAACATCGAGCTCCGCTGT	345		
QY	763	GAGCCTTGAGAGACTGCGCCGAGCTGTGTGCGGTCAAGTGTCTTGTGTGAGCGTGCCTCTGGCC	822		
Db	344	GAGCCTTGAGAGACTGCGCCGAGCTGTGTGCGGTCAAGTGTCTTGTGTGAGCGTGCCTCTGGCC	285		
QY	823	TCCGCAACCAAGACGTGTGCTGCTGTGACATGAGCTGTGCATGAGAAAGTGTGGCAAGTCT	882		
Db	284	TCCGCAACCAAGACGTGTGCTGCTGTGACATGAGCTGTGCATGAGAAAGTGTGGCAAGTCT	225		
QY	883	GTTGTGTGCCCAAGTGAATAATCCGGAAGGCACCATTTTAAACAATGAGATGTCAACGTGT	942		

```

Db      224 GTGGGGCCAAAGTAAATTCGGAGAGCAACATTTACATGACATGCTCAACGGTG 165
Qy      943 AAGGTGGGTGAGCCCAAGCCCTTCTCTGTAAGCATCTTTAATCTGTGGCAAGAG 1002
Db      164 AAGGTGGGTGA-CCCAAGGCTTCTCTGTAAGCATCTTTAATCTGTGGCAAGAG 106
Qy      1003 GTCCGTGTCACGTGTGAAGAGATGACACCATCATGAGAAATTGGTATATCATGGC 1062
Db      105 GTCCGTGTCACGTGTGAAGAGATGACACCATCATGAGAAATTGGTATATCATGGC 46
Qy      1063 AAAAAATCAAGTCTTAA 1080
Db      45 AAAAAATCAAGTCTTAA 28

RESULT 7
US-09-918-995-5220
; Sequence 5220, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5220
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(486)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-5220

Query Match      34.4%; Score 371.8; DB 10; Length 486;
Best Local Similarity 99.5%; Pred. No. 3.9e-109;
Matches 373; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AATCCGCTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGCGCAACCCGCTTC 60
Db      111 ATGCGCTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGCGCAACCCGCTTC 170
Qy      61 ATCATTTGCCGATGCGGCGCAACCAACAGGCGACCTGAGCGTAGCCAGGCGATATC 120
Db      171 ATCATTTGCCGATGCGGCGCAACCAACAGGCGACCTGAGCGTAGCCAGGCGATATC 230
Qy      121 CCGATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 180
Db      231 CCGATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 290
Qy      181 AAGTTAATCGAAAGACCTTTGGAGAGGCATACACTCGANACATTTCTGGGGGAAGAG 240
Db      291 AAGTTAATCGAAAGACCTTTGGAGAGGCATACACTCGANACATTTCTGGGGGAAGAG 350
Qy      241 TACGGGAGCAAAAGCAATCTGTAGTTCAAGCCATGACCAAGTACAGGAGCTGCAGAG 300
Db      351 TACGGGAGCAAAAGCAATCTGTAGTTCAAGCCATGACCAAGTACAGGAGCTGCAGAG 410
Qy      301 TACGGGAGCAAGGTGGATCTTCTCACTGCTCTGGCATGGATGATGAGTGA 360
Db      411 TACGGGAGCAAGGTGGATCTTCTCACTGCTCTGGCATGGATGATGAGTGA 470
Qy      361 TTCCTCATGACTG 375
Db      471 TTCCTCATGACTG 485

```

```

RESULT 8
US-09-918-995-28620
; Sequence 28620, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28620
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(442)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28620

Query Match      31.2%; Score 336.8; DB 10; Length 442;
Best Local Similarity 98.0%; Pred. No. 7.6e-98;
Matches 341; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      2 TGCCGTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGCGCAACCCGCTTCA 61
Db      90 TGCCGTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGCGCAACCCGCTTGA 149
Qy      62 TCATTGCCGAGATGGCCCAAGCCACCAAGGCGACCTTGAGCTGCAAGGCGATATCC 121
Db      150 TCATTGCCGAGATGGCCCAAGCCACCAAGGCGACCTTGAGCTGCAAGGCGATATCC 209
Qy      122 GCATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 181
Db      210 GCATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 269
Qy      182 AGTTAATCGAAAGACCTTTGGAGAGGCATACACTCGANACATTTCTGGGGGAAGAG 241
Db      270 AGTTAATCGAAAGACCTTTGGAGAGGCATACACTCGANACATTTCTGGGGGAAGAG 329
Qy      242 ACGGGAGCAAAAGCAATCTGTAGTTCAAGCCATGACCAAGTACAGGAGCTGCAGAG 301
Db      330 ACGGGAGCAAAAGCAATCTGTAGTTCAAGCCATGACCAAGTACAGGAGCTGCAGAG 389
Qy      302 ACGGGAGCAAGGTGGATCTTCTCACTGCTCTGGCATGGATGATGAGTGA 369
Db      390 ACGGGAGCAAGGTGGATCTTCTCACTGCTCTGGCATGGATGAGTGA 437

```

```

RESULT 9
US-09-918-995-25343
; Sequence 25343, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25343
; LENGTH: 417
; TYPE: DNA

```



```

APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA00591
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2649
LENGTH: 410
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (287)..(287)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (300)..(300)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (370)..(370)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (374)..(374)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (402)..(402)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (405)..(405)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (408)..(408)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2649

```

```

Query Match      23.7%; Score 256.2; DB 15; Length 410;
Best Local Similarity 97.2%; Pred. No. 8,2e-72;
Matches 280; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```

```

QY 548 AGGTACCGAGGATACCGGCTCCAGCTGAGAGGCAACCTGCGGTCATCTCGGAT 607
DB 110 AGGTACCGAGGATACCGGCTCCAGCTGAGAGGCAACCTGCGGTCATCTCGGAT 169
QY 608 ATCAGAGCTCTTCTGACATTCATAGGATTTCTGGGCTGAAACAGGATAGCA 667
DB 170 ATCAGAGCTCTTCTGACATTCATAGGATTTCTGGGCTGAAACAGGATAGCA 229
QY 668 TATCTGTGGCGAGGAGTGGCTCTGGGAGGCAAGGTGTTGAAAGCTCATTAATTGACA 727
DB 230 TATCTGTGGCGAGGAGTGGCTCTGGGAGGCAAGGTGTTGAAAGCTCATTAATTGACA 289
QY 728 AGACTGGAAGGGAGAGTACACTCGGCTCGCTGAGAGCTGGAAGACTGGCCG-AGCTG 786
DB 290 AGACTGGAAGGGAGAGTACACTCGGCTCGCTGAGAGCTGGAAGACTGGCCGAGAGCTG 349
QY 787 GTGGGTCATGTGCTGTGT-AGAGGCTGCGGCTGCGGCTCCCAAGCA 833
DB 350 GTGGGTCATGTGCTGTGTGTGAGAGGCTGCGGCTCCCAAGCA 397

```

```

RESULT 13
US-10-085-783A-31232
Sequence 31232, Application US/10085783A
Publication No. US2004003784A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.

```

```

TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 31232
LENGTH: 203
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (6)..(6)
OTHER INFORMATION: n is a, c, g, or t
NAME/KEY: misc feature
LOCATION: (31)..(31)
OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-31232

```

```

Query Match      16.7%; Score 180.4; DB 13; Length 203;
Best Local Similarity 98.4%; Pred. No. 1.7e-47;
Matches 181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 897 GAAATTCGGAAGGACCATTTCTACATGACATGCTACCGTGAAGGTGGGAGCC 956
DB 1 GAAATTCGGAAGGACCATTTCTACATGACATGCTACCGTGAAGGTGGGAGCC 60
QY 957 CAAGGCTATCTCTCTGAGGACATCTTATCTAGTGGGCAAGGCTCTGACTGT 1016
DB 61 CAAGGCTATCTCTCTGAGGACATCTTATCTAGTGGGCAAGGCTCTGACTGT 120
QY 1017 TGAAGAGATGACACCATCATGAGAAATTTGGTATATCATGCAAAAATTAATC 1076
DB 121 TGAAGAGATGACACCATCATGAGAAATTTGGTATATCATGCAAAAATTAATC 180
QY 1077 TTTA 1080
DB 181 TTTA 184

```

```

RESULT 14
US-10-242-535A-31232
Sequence 31232, Application US/10242535A
Publication No. US2004001363A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 31232
LENGTH: 203
TYPE: DNA
ORGANISM: Human
FEATURE:

```

NAME/KEY: misc feature  
 LOCATION: (6)..(6)  
 OTHER INFORMATION: n is a, c, g, or t  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (31)..(31)  
 OTHER INFORMATION: n is a, c, g, or t  
 US-10-242-535A-31232

Query Match 16.7%; Score 180.4; DB 16; Length 203;  
 Best Local Similarity 98.4%; Pred. No. 1.7e-47;  
 Matches 181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 897 GAAATCCCGAGGACCATCTTAACATGACATCATCAGCTGAGGTGGTGGCC 956  
 DB 1 GAAATCCCGAGGACCATCTTAACATGACATCATCAGCTGAGGTGGTGGCC 60  
 QY 957 CAAGCTATCTCTCTGAAGCATCTTTAATCTAGTGGCAAGAGTCTCTGCTACTGT 1016  
 DB 61 CAAGCTATCTCTCTGAAGCATCTTTAATCTAGTGGCAAGAGTCTCTGCTACTGT 120  
 QY 1017 TGAAGAGATGACACCATCATGAGAAATTGGTATATCATGACCAAAAAATCAAGTC 1076  
 DB 121 TGAAGAGATGACACCATCATGAGAAATTGGTATATCATGACCAAAAAATCAAGTC 180  
 QY 1077 TTAA 1080  
 DB 181 TTAA 184

RESULT 15  
 US-10-156-761-3360  
 Sequence 3360, Application US/10156761  
 Publication No. US20030119018A1  
 GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156, 761  
 PRIOR FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 3360  
 LENGTH: 933  
 TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(933)  
 US-10-156-761-3360

Query Match 12.5%; Score 135; DB 15; Length 933;  
 Best Local Similarity 49.3%; Pred. No. 2e-32;  
 Matches 413; Conservative 0; Mismatches 415; Indels 9; Gaps 2;

QY 40 GCGCGGCAACCCGCTTATCATCTGCGAGATCGGCCAGAAACCAACGAGGAGACTG 99  
 DB 43 GCGCGGCGCCACCCGCTTATCATCTGCGAGATCGGCCAGAAACCAACGAGGAGACTG 102  
 QY 100 GACGTAGCCAGGCGATATCCGATGCGCAAGAGATGTGGGGCTGATTGTGCCAAGTTC 159  
 DB 103 GAGAAAGCTTCAAGCTATGACGCGCCCGACAGCGCGCTGCGACGCGCTCAAGTTC 162  
 QY 160 CAGAAAGTGTAGTGAATTCAAGTTTATATGGAAGCTTGGAGAGCCATACACTTCG 219

DB 163 CAGAAAGCGACCCCGAGATCTGCAACCCCGCGGACAGATGGGACATGAGCGCGACACC 222  
 QY 220 AACCATCTCTGGGGAGAAAGCTGACGGGAGACCAAAACGACATCTGAGATTGACATGAC 279  
 DB 223 ---CCCTGGGGCGGAGTACCTTACATGACTGACCTGACCGGACCGGATGAGTGGAGAC 279  
 QY 280 CAGTACAGGAGCTGACAGAGTACCGCGGAGAGGTGGGATCTTCTCACTGCTTGGC 339  
 DB 280 GAGTACCGGACATGACAGAGTACCGCGGAGAGGTGGGATCTTCTCACTGCTTGGC 339  
 QY 340 ATGATGAGATGACAGTGAATTTCTGCACTGACATGATGTTCCATTTTCAAAGTTCAG 399  
 DB 340 TGGGACACCGAGCGCTCCCTCTTCTGAGAAAGTTCGACATCCCGCCCAAGGTGGCC 399  
 QY 400 TCTGAGACACTAATTAATTTCTTATCTGAAAAGACAGCCAAAAGGTGCGCCATG 459  
 DB 400 TCTGAGACACTAATTAATTTCTTATCTGAAAAGACAGCCAAAAGGTGCGCCATG 459  
 QY 460 GTGATCTCCAGTGGATGACATGATGACACACCATGAAAGATTATCAGATGTGAAG 519  
 DB 460 ATCTCTCCAGGCGATGTGAC-----CCGAAAGCATCCGTACCGGTGAGAGTTC 513  
 QY 520 CCCCTAACCCCACTTCTGCTTCTTGCAGTGTACAGGCGCATACCGGCTCCAGGCTGAG 579  
 DB 514 CTGGGCTCGGACAAATCTGCTCTTGCAGGCGCATCTGACGTACCGGCGAAAGGCGGAG 573  
 QY 580 GACGTCAACCTGCGGCTCATCTCGGAATATCAGAAAGCTTTCTGACATTTCCATAGGG 639  
 DB 574 GAGCTCAACCTCGGCTCATCTCGGAATATCAGAAAGCTTTCTGACATTTCCATAGGG 633  
 QY 640 TATTCTGGGATGAAACAGGATAGGATATCTGTGCGCGGACAGTGTCTTGGGGCGCAG 699  
 DB 634 TACTCCGGCCACAGAGCGGCTCGACAGACACGCTGCGCGGTGCTCCCTCGGCGCAG 693  
 QY 700 GTGTGGAAGTGCATTAATTGACCAAGACCTGGAAGGGAGTGAACCATGCGGCTCG 759  
 DB 694 TTCTGAGAGGCGCATCACTCGACCGGCGCATGTGGGGCTCCGACAGCGCGCTCC 753  
 QY 760 CTGAGCTCTGAGAACTGCGCGAGCTGTGTGCGGTCAGTGCTTTGTGAGCGTGCCTG 819  
 DB 754 GTGAGACCCGAGGCGCTCAAGCGCTGTGTGCGGTCAGTGCTTTGTGAGCGTGCCTG 813  
 QY 820 GGTCTCCCAACGAGAGCGTGTGCGCTGTGAGATGCGCTGCAATAGAGAGCTGGC 876  
 DB 814 GCGACGCGGTCAAGAGGCTTACAGAGTGTGAGCTGCGGCTCCATGAAGAGCTGGCG 870

Search completed: May 12, 2004, 12:37:27  
 Job time: 508.439 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:49 ; Search time 50.614 Seconds  
(without alignments)  
2004.184 Million cell updates/sec

Title: US-09-930-440B-6  
Perfect score: 1887  
Sequence: 1 MPLELELCPRKRWGQGHPCF.....EEDDTIMEELVNDHKKIKS 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	100.0	359	3 AAY96101	AAY96101 Human sla
2	1887	100.0	359	3 AAY90352	AAY90352 Human gly
3	1887	100.0	359	4 AAB84683	AAB84683 Amino aci
4	1887	100.0	359	4 AAO26546	AAO26546 Human SA
5	1883	99.8	359	5 ABB90221	ABB90221 Human pol
6	1880	99.6	359	4 AAM33986	AAM33986 Human pro
7	1876	99.4	359	4 AAB93183	AAB93183 Human pro
8	1789	94.8	400	4 AAM41772	AAM41772 Human pol
9	783	41.5	338	4 ABB71364	ABB71364 Drosophil
10	544.5	28.9	338	3 AAY68965	AAY68965 Cps2P pro
11	532.5	28.2	341	5 ABB26810	ABB26810 Streptococ
12	510.5	27.1	346	5 AAY96102	AAY96102 Escherich
13	510.5	27.1	346	4 AAB84684	AAB84684 Amino aci
14	510.5	27.1	346	6 AAO26547	AAO26547 Bacterial
15	419.5	22.2	125	4 ABB17251	ABB17251 Novel hum
16	395	20.9	153	4 ABB27650	ABB27650 Novel hum
17	382.5	20.3	349	5 AAM49716	AAM49716 N. mening
18	381	20.2	123	4 AAG75152	AAG75152 Human col
19	336.5	17.8	346	3 AAY97207	AAY97207 Campyloba
20	336.5	17.8	346	6 ABB18497	ABB18497 Campyloba
21	330.5	17.0	328	5 ABB78389	ABB78389 Amino aci
22	303	16.1	340	2 AAM98812	AAM98812 H. pylori
23	266.5	14.1	136	4 ABB27649	ABB27649 Novel hum
24	159.5	8.5	201	5 ABB51188	ABB51188 Helicobac
25	149.5	7.9	182	5 ABB50662	ABB50662 Helicobac

## ALIGNMENTS

26	113.5	6.0	333	6 ABB29572	ABB29572 Protein e
27	113.5	6.0	343	7 ADC97084	ADC97084 B. faeciu
28	110	5.8	265	4 AAB96139	AAB96139 Putative
29	108.5	5.7	633	6 ABB35582	ABB35582 Protein e
30	107.5	5.7	337	6 ABB23582	ABB23582 Protein e
31	107	5.7	330	6 ABB39673	ABB39673 Protein e
32	104.5	5.5	769	6 ABB70636	ABB70636 Phototrab
33	102.5	5.4	610	6 ABB06924	ABB06924 Altolococ
34	102.5	5.4	629	6 ABB06926	ABB06926 Altolococ
35	102.5	5.4	630	6 ABB06928	ABB06928 Altolococ
36	101.5	5.4	337	6 ABB25380	ABB25380 Protein e
37	101.5	5.4	341	6 ABB29304	ABB29304 Protein e
38	101	5.4	66	7 ADC79295	ADC79295 Z. elonga
39	101	5.4	350	7 ADC59279	ADC59279 Comamonas
40	101	5.4	350	7 ADC52503	ADC52503 Testoster
41	100	5.3	65	7 ADC79293	ADC79293 Z. elonga
42	100	5.3	547	6 ADA12178	ADA12178 Actinetera
43	100	5.3	547	7 ADC61133	ADC61133 Baeyer-Vi
44	100	5.3	547	7 ABB39407	ABB39407 Actinetera
45	100	5.3	1109	6 ABB71017	ABB71017 Human adi

## RESULT 1

AA96101	AA96101 standard; protein, 359 AA.
XX	AA96101;
AC	AA96101;
DT	19-DEC-2000 (first entry)
XX	
DE	Human sialic acid synthetase.
XX	
KM	Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;
KM	plasmidogen; transferin; thyrotropin; Na,K-ATPase.
XX	
OS	Homo sapiens.
XX	
PN	WC0200052135-A2.
XX	
PD	
XX	08-SEP-2000.
PF	01-MAR-2000; 2000MO-US005313.
XX	
PR	02-MAR-1999; 99US-0122582P.
PR	08-DEC-1999; 99US-0169624P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PA	(UYJO-) UNIV JOHN HOPKINS.
PA	(UYWY-) UNIV WYOMING.
XX	
PT	Benenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TR,
XX	WPI; 2000-572178/53.
DR	N-PSDB; AAA50569.
XX	
PT	Recombinant production of sialylated glycoproteins using cells in which
PT	the expression of enzymes, e.g. sialic acid synthetase, involved in the
PT	sialylation reaction has been altered.
XX	
PS	Claim 16; Page 105-106; 144pp; English.
XX	
CC	The present sequence is that of human sialic acid synthetase (SAS), an
CC	enzyme that condenses Mann6-P or Man-6-P with GDP to form Neu5Ac and
CC	RNA phosphates, respectively. The sequence was deduced from SAS cDNA (see
CC	AAA50569). Northern blots indicated ubiquitous transcription of the SAS
CC	gene in a selection of tissues. The invention provides methods and
CC	recombinantly engineered cells for producing glycoproteins having
CC	sialylated oligosaccharides. The methods involve altering the expression
CC	of enzymes involved in carbohydrate processing. A claimed cell producing
CC	sialylated glycoprotein at above endogenous levels expresses at least 1



(preferably human) enzyme selected from glucose-2 epimerase, an enzyme catalyzing the conversion of UDP-glucose to UDP-glucose-2, stialic acid synthetase, aldolase, CMP-stialic acid synthetase and CMP-stialic acid transporter at above endogenous levels. Endogenous N-acetylglucosaminidase activity may be suppressed. A claimed method for manipulating glycoprotein in an insect cell comprises enhancing the expression of 1 of the above enzymes, and a claimed method for producing stialylated glycoproteins involves expressing a heterologous protein (especially plasmidogen, transferrin, Na<sup>+</sup>/K<sup>+</sup>-ATPase or thyrotropin) in the insect cell. Yeast, insect, plant and bacterial host cells can be engineered to produce new forms of stialylated glycoproteins, higher concentrations of stialylated glycoproteins and/or elevated concentrations of donor substrates (e.g. nucleotide sugars) required for stialylation

Sequence 359 AA;

Query Match	100.0%	Score 1887	DB 3	Length 359
Best Local Similarity	100.0%	Pred. No. 9	4e-184	
Matches 359, Conservative	0	Mismatches	0	Indels 0, Gaps 0

QY 1 MRLBELCPGHWVGCHPCFLIAEIQNHQGLDVAKRMIRAKECGADCAFKQKSELEF 60  
DB 1 MRLBELCPGHWVGCHPCFLIAEIQNHQGLDVAKRMIRAKECGADCAFKQKSELEF 60

QY 61 KENRKALEPYTSKHSWGKTYGKHRLPESHQYREIQRVAEEVGIFPTASGMDENAVE 1200  
|||||  
61 KENRKALEPYTSKHSWGKTYGKHRLPESHQYREIQRVAEEVGIFPTASGMDENAVE 1200  
Db

QY 121 FLHELNVPPFKVSGDITNNPYLBJTAKKGRPMVSSGMSQSDTMMQYYQIVKPLNPNFC 180

Db 121 FLHELNVPPFKVSGDITNNPYLEKTAKKGRPMVSSGMSQSDTMMQYYQIVKPLNPNFC 180

QY	181	181	181
	FLOCTSAYPLOPEDVNLRVISEYOKLFPDIPICGSGHETGLAISVAALGAKVLERHIT		240
Db			
	FLOCTSAYPLOPEDVNLRVISEYOKLFPDIPICGSGHETGLAISVAALGAKVLERHIT		240

Qy 241 LDKTWKGS DHSASLEPGE LAI VR SVL VERALGS PTKÖLL PCMACNEKLGKSVAKVK 3000  
Db 241 LDKTWKGS DHSASLEPGE LAI VR SVL VERALGS PTKÖLL PCMACNEKLGKSVAKVK 3000

QY 301 IPEGTLTMMKLTVKVGEPRKVPREDIFNLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359  
301 |||||  
301 IPEGTLTMMKLTVKVGEPRKVPREDIFNLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359  
301 IPEGTLTMMKLTVKVGEPRKVPREDIFNLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359

RESULT 2  
 AAY90352  
 ID AAY90352 standard; protein; 359 AA.

AC	AAV90352;
XX	
DT	04-DEC-2000 (first entry)

DE Human glycosylation enzyme clone HA5A37 protein sequence

Human, glycosylation enzyme, glycolysis, myoglobinuria, tumor marker;  
 immunotherapy; cosmetic surgery, metabolism, immune system disorder;  
 haematopoietic cell deficiency, blood coagulation disorder, asthma;  
 afibrinogenemia, blood platelet disorder, thrombocytopenia, neoplasia;  
 autoimmune disorder, Addison's disease, multiple sclerosis, purpura;  
 allergic encephalomyelitis, allergic reaction, organ rejection;  
 graft-versus-host disease, inflammation, hyperproliferative disorder;  
 sacrocolitis, infection, gene therapy, CDP staltic acid synthetase.

05 Homo sapiens.

PN WO200052136-A2

PD 08-SEP-2000

PF 01-MAR-2000; 2000WO-US005325.

PR 02-MAR-1999; 99US-0122409P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Coleman TA;  
VV

DR WPI; 2000-572179/53  
DR N-2STB; AAA37763.

XX  
PT  
New human glycosylat

PT encoding the proteins for treating e.g., immune system disorders,

PT microbial diseases.

PS Claim 12; Page 110-111; 115pp; English.

CC This sequence represents a human glycosyltransferase clone of the  
CC invention, designated HASMA37. This protein clone is a sialic acid  
CC synthetase. The sequences are useful as reagents for the differential  
CC identification of the tissues or cell types present in a biological  
CC sample, as immunological probes, for treating a disease or condition  
CC resulting from under expression of such polypeptide, for the detection  
CC and/or treatment of disorders involving aberrant glycolysis, e.g. cramps  
CC myoglobinuria, and as tumour marker and/or immunotherapy targets. They  
CC may also be used to differentiate, proliferate and attract cells leading  
CC to the regeneration of tissues, to modulate mammalian characteristics  
CC (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism,  
CC anabolism processing utilisation and energy storage, to change a  
CC mammal's mental state by influencing and as a food additive or  
CC preservative. The proteins can be used to assay protein levels in a  
CC sample, as a marker or detector of an immune system disorder, to inhibit  
CC cytokine activity, and as a vaccine. They may further be used to treat  
CC immune system or of haemotopoietic cell deficiencies or disorders, blood  
CC coagulation disorders (e.g. afibrinogenaemia), blood platelet disorders  
CC (e.g. thrombocytopenia), wounds resulting from trauma or surgery,  
CC autoimmune disorders (e.g. Addison's disease, multiple sclerosis,  
CC allergic encephalomyelitis), allergic reactions (e.g. asthma), organ  
CC rejection, graft-versus-host disease, inflammation, hyperproliferative  
CC disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by  
CC viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include  
CC e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia,  
CC cellulitis), and diseases caused by parasites (e.g. amoebiasis,  
CC coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)

SQ Sequence 359 AA;

```

Query Match      100.0%   Score 1987;   DB 3;   Length 359;
Best Local Similarity 100.0%   Pred. No. 9 4e-184;
Matches 359;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```

1 MPELELCPRWVGQHCFTIAEIGQNHQGLDVAKRMIRAKECGADCAKFOKSELEF 60

Db 1 MPELELCPRWVGQHPCFIIAEIGQNHQGLDVAKRMIRAKECGADCAKFQKSELEF 60

61 KFNKALERTYTSKHSWGKTYGEHRLBFSHDQYRELORAAEEVCIFFTASGMDENAVE 120

Db 61 KFNKALERTYTSKISWGKTYGEHKRLLEFSHDQRELORYAEVGLFFIASGMDENAVE . 1

121 FHHELVNPFKVGSDINFPYLEKIAKGRMVISGQSMIDINQVIGIVNEINENC

DB 121 FHEELNVPFNVSBDLNNF FLEENALNORENVLDOONG.O.D.....

\_\_\_\_\_

100

\_\_\_\_\_

301 TDECTI TMDT TYNQZPBYA VBPEDT PNT YCKXLT VTVREDDT IMEEL VDNHGKKI XS 355

301 TPEGTITMDMLTVKVGEPKAYRPEDIENTLYGKKLVTVBEDDTIMEELVDNHGKKIKS 359

RESULT 3  
AAB84683  
ID AAB84683 standard; protein; 359 AA.  
XX  
AC AAB84683;  
XX  
DT 17-SEP-2001 (first entry)  
XX  
DE Amino acid sequence of a human sialic acid synthetase.  
XX  
KW Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;  
KW Cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;  
KW Sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;  
KW vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200142492-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 07-DEC-2000; 2000MO-US033136.  
XX  
PR 09-DEC-1999; 99US-0169839P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (UYJO ) UNIV JOHN HOPKINS.  
PA (UTEM ) UNIV TEMPLE.  
PA (UYWY-) UNIV WYOMING.  
XX  
PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;  
PI Jarvis D;  
XX  
DR MPI; 2001-441575/47.  
XX  
DR N-PSD3; AAB28458.  
XX  
PT Cells producing cytidine monophosphate-sialic acid and sialylated  
PT glycoprotein above endogenous levels for production of vaccines and  
PT therapeutics.  
XX  
PS Claim 16; Fig 32; 182pp; English.  
XX  
CC The specification describes a method for manipulating carbohydrate  
CC processing pathways in cells of interest. The methods are used to  
CC manipulate multiple pathways involved with the sialylation reaction by  
CC using recombinant DNA technology and substrate feeding approaches to  
CC enable the production of sialylated glycoproteins in the cells. The  
CC sialylation process involves the post-translational addition of the donor  
CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific  
CC acceptor carbohydrate. The cells express at least one enzyme, selected  
CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,  
CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The  
CC cells are useful for producing complex sialylated glycoproteins in cells  
CC of interest, especially insect cells. Glycoproteins containing sialylated  
CC oligosaccharides are useful as vaccines, therapeutics and diagnostic  
CC tools. Cells producing complex sialylated glycoproteins are useful for  
CC enhancing the value of heterologous expression systems and increasing the  
CC application of heterologous cell expression products as vaccines.  
CC Therapeutic and diagnostic tools as well as increasing the variety of  
CC heterologous proteins that can be produced and lowering biotechnology  
CC production costs. The present sequence represents a human sialic acid  
CC synthetase, which is used in the method of the invention  
XX  
SQ Sequence 359 AA;  
XX  
Query Match 100.0%; Score 1887; DB 4; Length 359;  
Best Local Similarity 100.0%; Pred. No. 9,4e-184;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MPLELELCPRWVGQHPFIIAIGQNHQDLDVAKMIRMAKCEGADCAKQKSELEF 60  
|||||

DB 1 MPLELELCPRWVGQHPFIIAIGQNHQDLDVAKMIRMAKCEGADCAKQKSELEF 60  
QY 61 KENRKALRPYTSKSHAGKTYGENHSHLEPSHDQYRELCQRYAEVGIPTTASGMDENAVE 120  
DB 61 KENRKALRPYTSKSHAGKTYGENHSHLEPSHDQYRELCQRYAEVGIPTTASGMDENAVE 120  
QY 121 FLHEINVPFPKVGSGDTNNPFLYEKTAKGPPMWTSSGMQSDMTKKQYQIVKPINPFC 180  
DB 121 FLHEINVPFPKVGSGDTNNPFLYEKTAKGPPMWTSSGMQSDMTKKQYQIVKPINPFC 180  
QY 161 FLOCTSAVPLQEDVNLAVISEYQCLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240  
DB 161 FLOCTSAVPLQEDVNLAVISEYQCLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240  
QY 241 LDKTWKGSDBHSAISLPEGLAELVRSVLVERALSGPTVQLLPCEMACNBKLGKSVAAVK 300  
DB 241 LDKTWKGSDBHSAISLPEGLAELVRSVLVERALSGPTVQLLPCEMACNBKLGKSVAAVK 300  
QY 301 IPEGTILMDMLTYVGEPRKAYPPEDIFNLVGKKVLYVREDDTMBELVDNHGKKITS 359  
DB 301 IPEGTILMDMLTYVGEPRKAYPPEDIFNLVGKKVLYVREDDTMBELVDNHGKKITS 359  
RESULT 4  
AA026546  
ID AA026546 standard; protein; 359 AA.  
XX  
AC AA026546;  
XX  
DT 06-MAR-2003 (first entry)  
XX  
DE Human SA synthetase protein.  
XX  
KW Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;  
KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAc; sialic acid;  
KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;  
KW transporter; sialylated glycoprotein; human.  
XX  
OS Homo sapiens.  
XX  
PN US2002142386-A1.  
XX  
PD 03-OCT-2002.  
XX  
PF 16-AUG-2001; 2001US-00930440.  
XX  
PR 02-MAR-1999; 99US-0122583P.  
PR 08-DEC-1999; 99US-0169624P.  
PR 25-AUG-2000; 2000US-0227579P.  
XX  
PA (BETE/) BETENBAUGH M J.  
PA (LAWR/) LAWRENCE S.  
PA (LEBY/) LEE Y C.  
PA (COLE/) COLEMAN T A.  
XX  
PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;  
XX  
DR MPI; 2003-102519/09.  
XX  
DR N-PSDB; AAL53993.  
XX  
PT Manipulating glycoprotein production in insect cell, involves enhancing  
PT expression of enzymes involved in carbohydrate processing pathway such as  
PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.  
XX  
PS Claim 16; Fig 32; 88pp; English.  
XX  
CC The invention relates to a novel method for manipulating glycoprotein  
CC production in an insect cell comprising enhancing expression of an  
CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one  
CC catalyzing conversion of UDP-GlcNAc to mannosyl (Man)NAc, sialic acid  
CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)  
CC synthetase or CMP-SA transporter, where the expression of each enzyme is  
CC enhanced to above endogenous levels. The novel method is useful for

manipulating glycoprotein production in an insect cell. Further methods of the invention are useful for producing sialylated glycoprotein. The sialylated glycoprotein produced by the above mentioned methods are useful as pharmaceutical compositions, vaccines, diagnostics and therapeutics. This sequence represents the human SA synthetase protein of the invention

Sequence 359 AA:

Query Match 100.0%; Score 1887; DB 6; Length 359;  
Best Local Similarity 100.0%; Pred. No. 9,4e-184;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLELELCGRWVGQGPCTIIAETIGNHQGLDVAKRMIMAKKCGADCAKFOKSELEF 60  
1 MLELELCGRWVGQGPCTIIAETIGNHQGLDVAKRMIMAKKCGADCAKFOKSELEF 60  
61 KFNKRALERPPTSKHSGKTYGKHLFESHDOYRELQRYAEVGIPTASGDMAYE 120  
61 KFNKRALERPPTSKHSGKTYGKHLFESHDOYRELQRYAEVGIPTASGDMAYE 120  
121 FLHELNVPPFKVSGDPTNNPPILEKTAAGKRPMTISSGQSMOTKQVQIVKPLNPF 180  
121 FLHELNVPPFKVSGDPTNNPPILEKTAAGKRPMTISSGQSMOTKQVQIVKPLNPF 180  
181 FLQCTSAVPLQPEVDNLRVISEYQKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240  
181 FLQCTSAVPLQPEVDNLRVISEYQKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240  
241 LDKTWKSDHSASLEPGLAEIVRSVLRALGSPTRKQLPCENACNEKLGKSVAAK 300  
241 LDKTWKSDHSASLEPGLAEIVRSVLRALGSPTRKQLPCENACNEKLGKSVAAK 300  
301 IPEGITLTMDMLTVKVGEPKAYPPEDIFNLGKKVLTVEEDDTIMEELVDNHGKKIKS 359  
301 IPEGITLTMDMLTVKVGEPKAYPPEDIFNLGKKVLTVEEDDTIMEELVDNHGKKIKS 359

RESULT 5  
ABB90221  
ID ABB90221 standard; protein; 359 AA.

XX ABB90221;

AC ABB90221;

XX 24-MAY-2002 (first entry)

DE Human polypeptide SEQ ID NO 2597.

Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antitubercular; hepatotropic; antidiabetic; antinflammatory; antitumor; vulvar; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX MO200190304-A2.

XX 29-NOV-2001.

XX 18-MAY-2001; 2001MO-US016450.

XX 19-MAY-2000; 2000US-02055152.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

XX N-PSDB; ABL90630.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive,

gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.

Claim 11; SEQ ID NO 2597; 2081bp + Sequence listing; English.

The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing infections; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WPI at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 359 AA:

Query Match 99.8%; Score 1883; DB 5; Length 359;  
Best Local Similarity 99.7%; Pred. No. 2,4e-183;  
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MLELELCGRWVGQGPCTIIAETIGNHQGLDVAKRMIMAKKCGADCAKFOKSELEF 60  
1 MLELELCGRWVGQGPCTIIAETIGNHQGLDVAKRMIMAKKCGADCAKFOKSELEF 60  
61 KFNKRALERPPTSKHSGKTYGKHLFESHDOYRELQRYAEVGIPTASGDMAYE 120  
61 KFNKRALERPPTSKHSGKTYGKHLFESHDOYRELQRYAEVGIPTASGDMAYE 120  
121 FLHELNVPPFKVSGDPTNNPPILEKTAAGKRPMTISSGQSMOTKQVQIVKPLNPF 180  
121 FLHELNVPPFKVSGDPTNNPPILEKTAAGKRPMTISSGQSMOTKQVQIVKPLNPF 180  
181 FLQCTSAVPLQPEVDNLRVISEYQKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240  
181 FLQCTSAVPLQPEVDNLRVISEYQKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240  
241 LDKTWKSDHSASLEPGLAEIVRSVLRALGSPTRKQLPCENACNEKLGKSVAAK 300  
241 LDKTWKSDHSASLEPGLAEIVRSVLRALGSPTRKQLPCENACNEKLGKSVAAK 300  
301 IPEGITLTMDMLTVKVGEPKAYPPEDIFNLGKKVLTVEEDDTIMEELVDNHGKKIKS 359  
301 IPEGITLTMDMLTVKVGEPKAYPPEDIFNLGKKVLTVEEDDTIMEELVDNHGKKIKS 359

RESULT 6  
AAM39986  
ID AAM39986 standard; protein; 359 AA.

XX AAM39986;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 3131.

Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

XX Homo sapiens.

PN MO200153312-A1.  
 XX 26-JUL-2001.  
 XX 26-DEC-2000; 2000MO-US034263.  
 XX 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00586042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX (HSE-) HSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Dmanac RT;  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AA159142.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX Example 4; SEQ ID NO 3131; 10078bp; English.  
 PS The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AA038642-AA042233) with neurotrophic,  
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Actin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening and  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 CC  
 XX Sequence 359 AA;  
 SQ  
 Query Match 99.4%; Score 1880; DB 4; Length 359;  
 Best Local Similarity 99.4%; Pred. No. 4.9e-183; Indels 0; Gaps 0;  
 Matches 357; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MPELELCGCRWVGCGHPCPIIAEIQNHQGDLDVAKRMIRAKKEGACDCAKQKSELEF 60  
 DB 1 MPELELCGCRWVGCGHPCPIIAEIQNHQGDLDVAKRMIRAKKEGACDCAKQKSELEF 60  
 QY 61 KERRKLEPRYTSKSHGKTYGKHRLRSHOYBELGYAREVGFPTASGMDMAVE 120  
 DB 61 KERRKLEPRYTSKSHGKTYGKHRLRSHOYBELGYAREVGFPTASGMDMAVE 120  
 QY 61 KERRKLEPRYTSKSHGKTYGKHRLRSHOYBELGYAREVGFPTASGMDMAVE 120  
 DB 61 KERRKLEPRYTSKSHGKTYGKHRLRSHOYBELGYAREVGFPTASGMDMAVE 120  
 QY 121 FLHEILNVPFVKVSGDTNNPFLYEKTAKKRPMVVISGMSQMDTKRQVQIVPLNPNFC 180  
 DB 121 FLHEILNVPFVKVSGDTNNPFLYEKTAKKRPMVVISGMSQMDTKRQVQIVPLNPNFC 180  
 QY 121 FLHEILNVPFVKVSGDTNNPFLYEKTAKKRPMVVISGMSQMDTKRQVQIVPLNPNFC 180  
 DB 121 FLHEILNVPFVKVSGDTNNPFLYEKTAKKRPMVVISGMSQMDTKRQVQIVPLNPNFC 180  
 QY 181 FLOCTSAVPEQEDVNLKRVSEYOKFPDIPICVSGHETSIASVAVALGAVALRRHT 240  
 DB 181 FLOCTSAVPEQEDVNLKRVSEYOKFPDIPICVSGHETSIASVAVALGAVALRRHT 240  
 QY 181 FLOCTSAVPEQEDVNLKRVSEYOKFPDIPICVSGHETSIASVAVALGAVALRRHT 240  
 DB 181 FLOCTSAVPEQEDVNLKRVSEYOKFPDIPICVSGHETSIASVAVALGAVALRRHT 240  
 QY 241 LDITWKGSDHSASLEPEGLAEIVRSVLRALGSPPTKOLLPECAKNEKLGKSVAAKYX 300  
 DB 241 LDITWKGSDHSASLEPEGLAEIVRSVLRALGSPPTKOLLPECAKNEKLGKSVAAKYX 300  
 QY 301 IPEGTILTMMLTVKGEPRKYPEDIFNLVGKKVLTVVEEDTILMEELVNHGKIKS 359  
 DB 301 IPEGTILTMMLTVKGEPRKYPEDIFNLVGKKVLTVVEEDTILMEELVNHGKIKS 359

DB 301 IPEGTILTMMLTVKGEPRKYPEDIFNLVGKKVLTVVEEDTILMEELVNHGKIKS 359  
 RESULT 7  
 ID AAB93183 standard; protein; 359 AA.  
 AC AAB93183;  
 XX 26-JUN-2001 (first entry)  
 DE Human protein sequence SEQ ID NO:12130.  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 PN EPI074617-A2.  
 PD 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-00116126.  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-0018776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.  
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 PI WPI; 2001-118749/34.  
 DR WPI; 2001-118749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 PS Claim 8; SEQ ID NO 12130; 2537bp + Sequence listing; English.  
 XX The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification; where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide complements at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX Sequence 359 AA;  
 SQ  
 Query Match 99.4%; Score 1876; DB 4; Length 359;  
 Best Local Similarity 99.2%; Pred. No. 1.3e-182; Indels 0; Gaps 0;  
 Matches 356; Conservative 1; Mismatches 2; Indels 0; Gaps 0;



PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 XX (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EM;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL15467.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 40884; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins (AB557737-  
 CC AB572072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)  
 CC  
 XX  
 SQ Sequence 338 AA;  
 XX  
 Query Match 41.5%; Score 783; DB 4; Length 338;  
 Best Local Similarity 45.6%; Pred. No. 6,5e-71;  
 Matches 154; Conservative 60; Mismatches 100; Indels 24; Gaps 3;  
 XX  
 QY 39 MIMAKGCGADCAKQSELEFKFNKRALBRYTSKSKGKTYGKHLBPSHDYREL 98  
 DB 1 MIMAKGAGCHCVFKSKDIPAKFTPSALDREYISDIAWKGTYGKHYLEFSKDYQL 60  
 QY 99 QRYAEVIGIFPTASGEMAVFEFLHNVPEFKYSGDINNFPLEKTKKGRPMVTS 158  
 DB 61 QAKKEINVDFTANMDEBSLEFLSALNVPFIKSGSDANNFLKAAANLPLVISTG 120  
 QY 159 MOSMDTKOVQYIVKPL-NENFCLOCTSAVPILOPEBVNLAVISEYKLEPPDIPGYS 217  
 DB 121 MQMTQVETRIYQVTRBSKEDYALMHCVSYPTDPKDCSLQILISVLRFPNVAIGYSG 180  
 QY 218 ETGALISVAALAAKTLBHTITLDTKMGSDHSAISPGELALVNSV----- 266  
 DB 181 ELGVITISQAAVLGARIIVERHFTIDKSKQSDHSCSELPQELKALTITATINPKLSV 240  
 QY 267 -RIVERALGS-----PTKQLPCENACNEKLGKSVAAVKIPEGTILTMDELTV 314  
 DB 241 PQELVYKLNGBDEHLEALQHVESKTIIPCELPCKNKGKSIIVAAARNINKGRIQLADMAI 300  
 QY 315 KVGEPKAVPEPDINLVGKKLVVVEEDDTIMEELVN 352  
 DB 301 KVSEPSGLTADFDLVGKELADNIGDEPILGNSIIN 338  
 XX  
 RESULT 10  
 AA68965  
 ID AA68965 standard; protein, 338 AA.  
 AC  
 AC AA68965;  
 XX  
 DE 30-MAY-2000 (first entry)  
 DE  
 DE Caps2P protein which is involved in sialic acid synthesis.  
 XX  
 XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;  
 XX capsular component; antigen; regulation; chain length determination;  
 XX complement-mediated opsonophagocytosis; serotype-specific detection;  
 XX antigen; vaccine; Streptococcal disease; ORF 22; ORF 25; ORF 26; ORF 27;  
 XX Caps2B; Caps2C; Caps2D; Caps2E; Caps2F; Caps2G; Caps2H; Caps2I; Caps2J; Caps2K;

KM Caps2O; Caps2P; Caps2Q; Caps2R; Caps2S; Caps2T.  
 OS Streptococcus suis.  
 XX  
 XX MO200005378-A2.  
 XX  
 XX 03-FEB-2000.  
 XX  
 XX 19-JUL-1999; 99MO-NL000460.  
 XX  
 XX 22-JUL-1998; 98EP-00202465.  
 PR 22-JUL-1998; 98EP-00202467.  
 XX  
 XX (DIEN-) STICHTING DIENST LANDBOUDWONKINDIG ONDERZOE.  
 XX  
 XX Smith HR;  
 PI  
 DR WPI; 2000-195104/17.  
 DR N-PSDB; AA260929.  
 XX  
 XX New nucleic acid containing the capsular gene cluster of *Streptococcus*  
 PT suis, used for serotype-specific detection and to generate antigens or  
 PT mutants for vaccination.  
 XX  
 XX  
 PS Disclosure; Fig 3; 14pp; English.  
 XX  
 CC The proteins AA68950-69 are encoded by the capsular gene cluster of  
 CC *Streptococcus suis* serotype 2. The genes in this cluster are involved in  
 CC polysaccharide biosynthesis of capsular components and antigens. The  
 CC proteins are involved in regulation (CpsA), chain length determination  
 CC (CpsB, CpsC), export (CpsD), and biosynthesis (CpsE, CpsF, CpsG, CpsH,  
 CC CpsI, CpsK). The capsule confers bacterium resistance to complement-  
 CC mediated opsonophagocytosis. The gene cluster is used as a source of  
 CC probes and primers for serotype-specific detection of *S. suis* and is also  
 CC useful for recombinant production of the proteins. The proteins are then  
 CC useful for producing antigens that can be used in vaccines, for  
 CC controlling or eradicating a Streptococcal disease, in humans or animals,  
 CC e.g. against *S. suis* in pigs  
 CC  
 XX  
 SQ Sequence 338 AA;  
 XX  
 Query Match 28.9%; Score 544.5; DB 3; Length 338;  
 Best Local Similarity 36.0%; Pred. No. 1.6e-46;  
 Matches 123; Conservative 73; Mismatches 129; Indels 17; Gaps 8;  
 XX  
 QY 20 FTIABIGONHGDLDYAKMIRAKCGADCAKQSELEFKFNKRALBRYTSKSKGK 79  
 DB 3 YIIABIGONHGDYHAKKRVAVAVDCGDVAFQCFKADLLISKAPAAEY-QKITTEG 61  
 QY 80 TYG-EHKHLEFSDQYRELQRYAEVGIFFTASGMDENAVFEFLHNVPEFKYSG 137  
 DB 62 SDQLEWTRLELSEFSEYIDLRYCLEKGVDPSTFDESDLPFLISTDPVVKIISGEL 121  
 QY 138 NNEPVYEKTRKGRPMVVISGMSQNDTMKOVQYIVPLNN-----PCFLOCTSAVPI 193  
 DB 122 TNLPYEKGRQKKKYLISGAVMD---RHHAVKILQENGTIDISILCTEYRPPR 178  
 QY 194 DVNIRVISEYOKLPDIPIGYSGHETGIALISVAALGAKVLRHTITLDTKMGSDH 253  
 DB 179 ALNINVLHTLKEFPNLTIGSDHSVSEVPVIAAAMGAEILIKHFTLDNEMSGPDH 238  
 QY 254 LEPGELAEIVASRYIVVERALGSPITKQLPCENACNEKLGKSVAAVKIPEGTIL 311  
 DB 239 AFDPIILAVKGVRIYVQSLKTEKE--PEEVARKIVARKSIIVAKKALAKSEVTEEN 296  
 QY 312 LTVKVGEP-KAYPPEDIFNLVGGKVLVVEEDDTIMEELVN 352  
 DB 297 ITVK--RPGGISPMEMYKVLGVSEODPREDQNIHSAFEN 336  
 XX  
 RESULT 11  
 ABP26810  
 ID ABP26810 standard; protein, 341 AA.

XX ABP26810;  
AC  
XX  
XX 02-JUL-2002 (first entry)  
DT  
XX  
XX Streptococcus polypeptide SEQ ID NO 2796.  
DE  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX anti-inflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
XX Streptococcus agalactiae.  
OS  
XX WO200234771-A2.  
XX  
XX 02-MAY-2002.  
XX  
XX 29-OCT-2001; 2001WO-GB004789.  
XX  
XX 27-OCT-2000; 2000GB-00026333.  
XX 24-NOV-2000; 2000GB-00028727.  
XX 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Maignant V, Margatit Y Rost, Grandi G, Fraser C;  
XX Telford H;  
XX  
XX WPI; 2002-352536/38.  
XX N-PSDB; ABN67441.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
XX disease caused by Streptococcus bacteria, such as meningitis, and for  
XX detecting a compound that binds to the protein.  
XX  
XX Claim 1, Page 3430; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
XX the specification. The proteins have antibacterial and anti-inflammatory  
XX activity. (I), nucleic acids encoding (II), ABN6044-ABN71526 and  
XX antibodies that bind (I) are used in the manufacture of medicaments for  
XX the treatment or prevention of infection or disease caused by  
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX biological sample. (I) is used to determine whether a compound binds to  
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX used as a vaccine or diagnostic composition. The disease caused by  
XX Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX acid encoding (I) may be used to recombinantly produce (I) and may be  
XX used in gene therapy. Antibodies to (I) are used for affinity  
XX chromatography, immunoassays, and distinguishing/identifying  
XX Streptococcus proteins  
XX  
XX Sequence 341 AA;  
SQ

Query Match 28.2%; Score 532.5; DB 5; length 341;  
Best Local Similarity 36.8%; Pred. No. 2.7e-95;  
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FIIAIGCHQHDIDVAKMIRAKGCGADKQKSEIEFNRKALRPTSKHSMXK 79  
DB 3 YIIAIGCHQHDIDVAKMIRAKGCGADKQKSEIEFNRKALRPTSKHSMXK 61  
QY 80 TVG--EAKHLESHDQYRZELQYAEVGIFFTAQMDMAVEFLHNLVPEFKVSGSDT 137  
DB 62 ASQGLMTRELSFEYILMDRYALISKVETSTPDESELEFLISTMPYKIPSGRI 121  
QY 138 NNPPTLEKAKKRPVWISSGQSMOTMGOVQIVKGLPN---PCFQCSAYLQPE 193  
DB 122 TMLPYLEKIGKQKQKVLSTGMA--VMEIHQAVNLKONGTDTISILHCTETPTPTP 178

QY 194 DVNLTIVSEYQKLEFPDIPIGYSGHETGIAISVAALGAKTLERHITLDTKWSDSAS 253  
DB 179 SLINLVHTLMDREFKDLITIGSDHSISEVPIAAAGAEVIEHFTLDTMBSGDHAS 238  
QY 254 LEPGELAEVRSVLEVERALGSPFKOILPCMACNENKIGKSVAKV1PEGTIITMDMLT 313  
DB 239 ATPDILALVKGAVIVGALGRPEKIPDPVEERKRIYARSSVVALPKIKGDIVSIEMIT 298  
QY 314 VKGEP-KAPPEDIFNIVGKVLVTVBEDDTIMEELVDN 352  
DB 299 VK--RPGNGISPMWYDILGQEAQDDPEDEEVIDRSRPN 336

RESULT 12  
ID AA96102 standard; protein, 346 AA.  
AA96102;  
AC  
XX  
XX 19-DEC-2000 (first entry)  
DT  
XX  
XX Escherichia coli sialic acid synthetase.  
DE  
XX Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;  
XX plasmidogen; transferrin; thyrotropin; Na<sup>+</sup>, K<sup>+</sup>-ATPase.  
XX  
XX Escherichia coli.  
XX  
XX WO200052135-A2.  
XX  
XX 08-SEP-2000.  
XX  
XX 01-MAR-2000; 2000WO-US005313.  
XX PF  
XX 02-MAR-1999; 99US-0122582P.  
XX 08-DEC-1999; 99US-0169624P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (UYO) UNIT JOHNS HOPKINS.  
XX (UWY-) UNIT WYOMING.  
XX  
XX Betenbough MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;  
XX WPI; 2000-572178/53.  
XX N-PSDB; AAA50570.  
XX  
XX Recombinant production of sialylated glycoproteins using cells in which  
XX the expression of enzymes, e.g. sialic acid synthetase, involved in the  
XX sialylation reaction has been altered.  
XX  
XX Example 6, Page 108; 144pp; English.  
XX  
XX The present sequence is that of Escherichia coli sialic acid synthetase,  
XX encoded by the nub gene (see AAA50570). The nub gene was used to  
XX isolate the corresponding human gene (see AAA50569) for sialic acid  
XX synthetase (see AA96101). The invention provides methods and  
XX recombinantly engineered cells for producing glycoproteins having  
XX sialylated oligosaccharides. The methods involve altering the expression  
XX of enzymes involved in carbohydrate processing. A claimed cell producing  
XX sialylated glycoprotein at above endogenous levels expresses at least 1  
XX (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme  
XX catalyzing the conversion of UDP-GlcNAc to ManNAc, sialic acid  
XX synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid  
XX transporter at above endogenous levels. A claimed method for manipulating  
XX glycoprotein in an insect cell comprises enhancing the expression of 1 of  
XX the above enzymes, and a claimed method for producing sialylated  
XX glycoproteins involves expressing a heterologous protein (especially  
XX plasmidogen, transferrin, Na<sup>+</sup>, K<sup>+</sup>-ATPase or thyrotropin) in the insect  
XX cell. Yeast, insect, fungal, plant and bacterial host cells can be  
XX engineered to produce new forms of sialylated glycoproteins, higher  
XX concentrations of sialylated glycoproteins and/or elevated concentrations  
XX of donor substrates (e.g. nucleotide sugars) required for sialylation

Query Match	27.1%	Score 510.5	DB 3	Length 346
Best Local Similarity	36.3%	Pred. No. 4,9e-43		
Matches 123	Conservative 61	Mismatches 140	Indels 15	Gaps 6
QY	20	FLIAEIGNHOGDLDVAAKMTIMAECCADCAKPEKSEIEFENRKALERPYTSKSWG-78		
DB	5	YVAIEGNNHNGSVDIAREMITKAEAGVNAVKPFTPADKDISAIDRAEYQIKNTGEL-64		
QY	79	KTYGHHKALPESHQDQRELCRYAAEEVGIFTLASGMDENAEFFHELAVNPPFKVSSGDTN-138		
DB	65	ESQLEMTKKLEMKYDYLHLMAYVASLINDVSTFPEDESDIDFLASKOKIKWIKISGELL-124		
QY	139	NFPYLTKAKKGRP---NWISSGMSMDTMKVQYIV---KPLNPFGLQCTSAYPELP-192		
DB	125	NLPYIEKIAKLPDPKIIIIISTGMFTIDEIKOSVSLFINNKPVGENTIIILHNTYPTFP-184		
QY	193	EDVNLRVISEYOKLPPDIPIGYSGHETGIAISVAALGAKYLERHITLDTKWSGDSA-252		
DB	185	EDVNLRVISEYOKLPPDIPIGYSGHETGIAISVAALGAKYLERHITLDTKWSGDSA-244		
QY	253	SLEPELAEVRSVNLVERALGSPFKQILLPCMACNEKLGKSVNAKVIPEGTIIITMDL-312		
DB	245	SLEPELAEVRSVNLVERALGSPFKQILLPCMACNEKLGKSVNAKVIPEGTIIITMDL-304		
QY	313	TVKVGEP-KAYPPEDEFNLVGKAVLWVEEDDTIMEELV-350		
DB	305	TTK--RPGNGISPKMEWNTLGR-----IAGQFIIDELL-336		
RESULT 13				
AA84684				
AA84684 standard; protein; 346 AA.				
AA84684:				
17-SEP-2001 (first entry)				
Amino acid sequence of a bacterial sialic acid synthetase.				
Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;				
cyclidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;				
sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;				
vaccine.				
Escherichia coli.				
WO200142492-A1.				
14-JUN-2001.				
07-DEC-2000; 2000WO-US03136.				
09-DEC-1999; 99US-0169839P.				
(HUMA-) HUMAN GENOME SCI INC.				
(UYCO) UNIV JOHNS HOPKINS.				
(UTEM) UNIV TEMPLE.				
(UYMY-) UNIV WYOMING.				
Berenbough MJ, Lawrence S, Lee YC, Coleman TA, Palter K;				
Jarvis D;				
WPI; 2001-441575/47.				
N-PDB; AAH28459.				
Cells producing cytidine monophosphate-sialic acid and sialylated				
glycoprotein above endogenous levels for production of vaccines and				
therapeutics.				
Example 5; Page 165-166; 182pp; English.				

[illegible]



XX 16-AUG-2001; 2001US-00930440.  
 PF 02-MAR-1999; 99US-0122582P.  
 XX 08-DEC-1999; 99US-0169624P.  
 PR 25-AUG-2000; 2000US-0227579P.  
 XX (BRET/) BETENBAUGH M J.  
 PA (LAWR/) LAWRENCE S.  
 PA (LEBY/) LEE Y C.  
 PA (COLE/) COLEMAN T A.  
 XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;  
 PI WPI: 2003-102519/09.  
 DR N-PSDB; AAL53994.  
 XX Manipulating glycoprotein production in insect cell, involves enhancing  
 PT expression of enzymes involved in carbohydrate processing pathway such as  
 PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.  
 XX Disclosure; Fig 35D; 86pp; English.  
 PS The invention relates to a novel method for manipulating glycoprotein  
 CC production in an insect cell comprising enhancing expression of an  
 CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one  
 CC catalyzing conversion of UDP-GlcNAc to mannose (Man)NAc, sialic acid  
 CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)  
 CC synthetase or CMP-SA transporter, where the expression of each enzyme is  
 CC enhanced to above endogenous levels. The novel method is useful for  
 CC manipulating glycoprotein production in an insect cell. Further methods  
 CC of the invention are useful for producing sialylated glycoprotein. The  
 CC sialylated glycoprotein produced by the above mentioned methods are  
 CC useful as pharmaceutical compositions, vaccines, diagnostics and  
 CC therapeutics. This sequence represents the bacterial sialic acid (NeuB)  
 CC synthetase protein of the invention  
 XX  
 XX Sequence 346 AA;  
 SQ  
 Query Match 27.1%; Score 510.5; DB 6; Length 346;  
 Best Local Similarity 36.3%; Pred. No. 4.9e-43;  
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;  
 QY 20 FIIAIGNHQGDLDVAKRMIMAKECGADCKAFKPSXSEIEFFNKKALERTPTSGHSG- 78  
 DB 5 YIVARIQCHNGSVIARFMILKAEAGVNAVAFQTFKADKLISAIAPAEYQINTEEL 64  
 QY 79 KTYGEHKRLERSHOYRELQYAEVGIFFTASQMDENAVFLHELNVPPFKVSGSDTN 138  
 DB 65 BQOLETKLEMKDYIHLMEYANSINDVSTFDESDIDFLASLQKIMKISGELL 124  
 QY 139 NFPLYEKTKAKRP--WVSSGMSMDTKOVQIV--KPLNFNFCLOCTSAIYLP 192  
 DB 125 NIPYLEKIKLPIPKKIIISTGMATIDIKQSVSIFINNKKPVGVGITILHNCITPTPPF 184  
 QY 193 EDVNLKVISYOKLPFDIPIGVSGETGIAISVAVNALGAKVLEPHITLDTKWSQSDSA 252  
 DB 185 EDVNLNAILMDLKHFFPKNIGFSDHSGTYAIAVAPYITTEKGFILDKMSAPDHLA 244  
 QY 253 SLEPGLAELVRSYVLVERALGSPYKQLPCMAQNEKLGKSVAKVPIEGTILTMML 312  
 DB 245 SLEPDELKHLGIGVCEKSLGNSKRVYASERKNKIVARKSIIAKTEIKKEVSEKRI 304  
 QY 313 TVKVGEP-KAYPEDIENLVGKKVLTVEEDDTIMEIV 350  
 DB 305 TTK--RPGNGISPEKMTNLGK-----IADDFIDDELI 336

RESULT 15  
 ABG17251  
 ID ABG17251 standard; protein; 125 AA.  
 XX  
 AC ABG17251;

XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #17242.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX Human; supplement; medical imaging; diagnostic; genetic disorder.  
 KW Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US008631.  
 PE 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Dmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB; AAS81438.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 47610; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptides and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 XX  
 XX Sequence 125 AA;  
 SQ  
 Query Match 22.2%; Score 419.5; DB 4; Length 125;  
 Best Local Similarity 79.2%; Pred. No. 2e-34;  
 Matches 80; Conservative 6; Mismatches 14; Indels 1; Gaps 1;  
 QY 1 MFLBELCPGRWVGQHPFCFIIAIGNHQGDLDVAKRMIMAKECGADCKAFKPSXSEIEF 60  
 DB 9 MPVELIELCPGRWVGQHPFCFIIAIGNHQGDLDVAKRMIMAKECGADCKAFKPSXSEIEF 68  
 QY 61 KPNKALERTPTSGHSGK--TYGEHKRLERSHOYRELQ 100  
 DB 69 KPNKALERTPTSGHSGSDYRGAQTDIKFNHQQVGAAR 109

Search completed: May 6, 2004, 09:04:23  
 Job time : 54.614 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

# OM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 13.6362 Seconds  
(without alignments)  
2532.427 Million cell updates/sec

Title: US-09-930-440B-6

Perfect score: 1887

Sequence: 1 MPELELCGRWVGQHPGF.....EEDDTIMEELVDNHGKIKS 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	1804	95.6	359	2	JC7321
2	619	32.8	337	2	H64432
3	532.5	28.2	341	2	T44651
4	510.5	27.1	346	2	I69836
5	487	25.8	334	2	D81276
6	419	22.2	350	2	F97169
7	382.5	20.3	349	2	S60760
8	370	19.6	373	2	S39722
9	358	19.0	356	2	D87604
10	356.5	18.9	343	2	B81375
11	354	18.8	343	2	H81318
12	303	16.1	340	2	B64542
13	287	15.2	340	2	B71965
14	226	12.0	378	2	H71307
15	122.5	6.5	331	2	B90173
16	119	6.3	332	2	A11899
17	110	5.8	134	2	S53514
18	110	5.8	125	2	B75161
19	108.5	5.7	351	2	S74729
20	108.5	5.7	633	2	D64222
21	108	5.7	63	2	S04974
22	108	5.7	338	2	S04388
23	107.5	5.7	337	2	A97010
24	107	5.7	276	2	B72643
25	106.5	5.6	478	2	B70342
26	102.5	5.4	1140	2	S73786
27	102	5.4	63	2	S04973
28	101	5.4	1107	1	S52517
29	101	5.4	1175	2	T20346

30	100.5	5.3	481	2	H69593	6-phospho-beta-glu
31	100	5.3	1109	2	S53601	myosin-1C [similar
32	98.5	5.2	371	2	A39198	SASP degradation s
33	96.5	5.1	454	2	A43501	sucrose-6-phosphat
34	96.5	5.1	1356	2	T16754	hypothetical prote
35	96	5.1	880	2	F71652	pyruvate, phosphate
36	95.5	5.1	513	2	A05029	reticulocyte-bindl
37	95.5	5.1	2829	2	A42771	reticulocyte-bindl
38	94	5.0	64	1	FDR1RE	antifreeze protein
39	94	5.0	312	2	D90459	phosphoenolpyruvat
40	94	5.0	435	2	G96431	conserved hypothet
41	93.5	5.0	329	2	H83099	probable oxidoredu
42	93	4.9	64	2	S53512	antifreeze protein
43	93	4.9	542	2	A28550	cyclohexanone mono
44	92.5	4.9	642	2	D90558	topoisomerase iv s
45	92.5	4.9	1258	2	T29041	hypothetical prote

## ALIGNMENTS

### RESULT 1

JC7321

N-acetylneuraminic acid 9-phosphate synthase (EC 4.1.3.-) - mouse

N/Alternate names: N-acetylneuraminic acid synthase homolog

C/Spectrum: Mus musculus (house mouse)

C/Date: 08-Sep-2000 #sequence revision 08-Sep-2000 #text\_change 06-Oct-2000

C/Accession: JC7321

R/Nakata, D.; Close, B.E.; Colley, K.J.; Matsuda, T.; Kitajima, K.

Biochem. Biophys. Res. Commun. 273, 642-648, 2000

A/Title: Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosph

A/Reference number: JC7321

A/Accession: JC7321

A/Molecule type: mRNA

A/Residues: 1-359 <NA>

A/Cross-references: DDBJ:AB041263

C/Comment: This enzyme is a cytosolic enzyme and as a housekeeping enzyme as well, whic

C/Genetics:

A/Name: Sialic acid

A/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpi

C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 95.6%; Score 1804; DB 2; Length 359;

Best Local Similarity 94.2%; Pred. No. 1e-133;

Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY	1	MPELELCGRWVGQHPGFIIAETGQNHQGLDVAKMIRAKGADCAFOKSELEF	60
DB	1	MPELELCGRWVGQHPGFIIAETGQNHQGLDVAKMIRAKGADCAFOKSELEF	60
QY	61	KNNKRLSRPYTSKSGMGTYGEHKKLEFSDQYRELQYVAEYGIPTFASGEMAVE	120
DB	61	KNNKRLSRPYTSKSGMGTYGEHKKLEFSDQYRELQYVAEYGIPTFASGEMAVE	120
QY	121	FLHELVNPEFKVSGSDTNNPYLEKTAKKGRPVISGQSGMDTKQYQVYKPLNPNFC	180
DB	121	FLHELVNPEFKVSGSDTNNPYLEKTAKKGRPVISGQSGMDTKQYQVYKPLNPNFC	180
QY	181	FIQCTSAVLOPEDNMLRVISEYQKLPDIPIGSGHETGINSVAVALGAKVLERHT	240
DB	181	FIQCTSAVLOPEDNMLRVISEYQKLPDIPIGSGHETGINSVAVALGAKVLERHT	240
QY	241	LDKTKGSDHSASLPEGEIAELVRSVRLVERALGSPPTQQLPCENACNEKLGKSVAYKX	300
DB	241	LDKTKGSDHSASLPEGEIAELVRSVRLVERALGSPPTQQLPCENACNEKLGKSVAYKX	300
QY	301	IEGTLTMDMLTVYVGEKAPPPEDINLVCKYLVNVEEDTIMEELVDNHGKIKS	359
DB	301	IEGTLTMDMLTVYVGEKAPPPEDINLVCKYLVNVEEDTIMEELVDNHGKIKS	359



N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1327 [imported] - Campylobacter jejuni  
 C:Species: Campylobacter jejuni  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: D81276  
 R:Perkhill, J.; Wren, B.W.; Mungall, K.; Kerley, J.M.; Churcher, C.; Basham, D.; Chilling  
 Nure, A.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVlier, A.; Whitehead, S.; Barral  
 C.M.; 403, 665-668, 2000  
 A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
 A:Reference number: A81250, PMID:20150312; PMID:10688204  
 A:Accession: D81276  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1334 <PAR>  
 A:Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:CA87354.1; PID:9696876  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: neuB2; Cj1327  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 25.8%; Score 487; DB 2; Length 334;  
 Best Local Similarity 35.1%; Pred. No. 3.6e-30;  
 Matches 118; Conservative 64; Mismatches 138; Indels 16; Gaps 5;

QY 21 IIAEIGNHGGDDVAKRMIMAKGADCAKFKSELEFNRKALREPYTSKSWG-K 79  
 DB 6 IIAEAGVNHGGDDVAKRMIMAKGADCAKFKSELEFNRKALREPYTSKSWG-K 65  
 QY 80 TYGHEKHELEFSDQYELQRYAEVGIFFTASGDMAVEFLHELVNPFVKVSGSDTN 139  
 DB 66 SQQWVQKLELDKAKHELEIHAKKCNIAELSTPDLSEVDLNEGLKFKIPSGEITN 125  
 QY 140 FPIYLEKAKGKRPVIVSSGQSMQDTKQVQIV---KPLNPNCFQCTSAVYLOGEDN 136  
 DB 126 LPIYKIKAKNKKITLSTGANLGEIEBALNVCCKGAKQNTLMLCTETYPAPFENVN 185  
 QY 197 LRVISEYQKLPDPIPIGSGHEGTGIALSVAAVAGATVETRHITLDTKWSGSDHSLSP 256  
 DB 186 LKMQSLKAP-KLDVGVSHTRGIRHISLAVALGACVIEKHPIFLDRMGSPHKSLEP 244  
 QY 257 GELAEVRSVRLVERALGSPKQLPCEMACNEGLGSVYAKIPEGTILMDLTVKY 316  
 DB 245 QELMLCTQIQIQOKAMGDKKASKGQKNNINIVRSIVAKDKKGEIFSGNLTTR- 303  
 QY 317 GEPKAYPEPDI-----NVEGKVLTVPEEDTIME 347  
 DB 304 -----RANGISMRVPEFLGKATATYKEDDLIRE 334

RESULT 6  
 F97169  
 A:Title: Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: F97169  
 R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J.; Daily, M.J.; Bennett, G.N.; Kocin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
 A:Reference number: A96900; PMID:21359325; PMID:21359325  
 A:Accession: F97169  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-350 <KTR>  
 A:Cross-references: GB:AE001437; PIDN:AAK80145.1; PID:G15025183; GSPDB:GNO0168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC2187  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 22.2%; Score 419; DB 2; Length 350;  
 Best Local Similarity 32.3%; Pred. No. 3.6e-25;  
 Matches 112; Conservative 73; Mismatches 132; Indels 30; Gaps 10;

QY 10 GEMVGGHPCFIILAIQNHQEDLVAKRMIMAKGADCAKFK-----KSELEF 60  
 DB 9 GKIKGGGRTFIILAEASNAHNDPDRAVEIILAKRSGDAKILQYTPDITTPSDNAY 68  
 QY 61 KENRKALREPYTSKSWGTYGHEKHELEFSDQYR-ELQRYAEVGIFFTASGDMEMAV 119  
 DB 69 POLKQCTITWDGTLH---KLVE-----AYTWQOPKLEIAEBGLICFSSPNTSV 120  
 QY 120 EFLAEVLNVPFFVSGSDTNPPYLEKTKAKGKRPVIVSSGQSMQDTKQVQIV---KPL 175  
 DB 121 DFLKKEKVPAYKVASFELTDIPFIEYIAKGRKVINSTGIAEMG---EYEAVAACRQ 177  
 QY 176 NENPCLQCTSAVPIQPEVNLRVISEVQKLPDPIPIGSGHEGTGIALSVAAVAGATKL 235  
 DB 178 NENVILKCSSYSPPELEINIKTIPNNKEAF-NCVGLSDHTMGVATPAVALDAIYI 236  
 QY 236 EKHITLDTKWSGSDHSLSPGELAEVRSVRLVERALGSPKQLPCEMACNEKLGKGV 295  
 DB 237 EKHFTLKSDGDPDSAPSMKEPEFSAMVKSIREVERALGKTYELTE-KQNSRQSRSL 295  
 QY 296 VAKWKIPBGTITLMDLTVKVEEPK-AIPPEIDPVLVKKVLYTVBE 341  
 DB 296 FVVKDKIKKGETPTKX--NVKXIRPAPGLTKYIEVIGKRAVDVVK 340

RESULT 7  
 S60760  
 A:Title: Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 19-Jan-2001  
 C:Accession: S60760; B81241  
 R:Richards, U.; Mueller, A.; Hamerschmidt, S.; Gerardy-Schahn, R.; Frosch, M.  
 Mol. Microbiol. 14, 141-149, 1994  
 A>Title: Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid  
 A:Reference number: S60758; PMID:9531727; PMID:783052  
 A:Accession: S60760  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-349 <EDM>  
 A:Cross-references: EMBL:M95053; NID:9520732; PIDN:AA420477.1; PID:9520735  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994  
 R:Petelin, H.; Saunders, M.V.; Heideberg, J.; Jeffries, A.C.; Nelson, K.B.; Bisen, J.;  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 H.; Qin, H.; Vanatavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.  
 Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; V  
 A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: A81000; PMID:20175755; PMID:10710307  
 A:Accession: B81241  
 A:Molecule type: DNA  
 A:Residues: 1-349 <DET>  
 A:Cross-references: GB:AE002366; GB:AE002098; NID:97225284; PIDN:AAF40535.1; PID:9722528  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB0066  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 20.3%; Score 382.5; DB 2; Length 349;  
 Best Local Similarity 30.5%; Pred. No. 2.6e-22;  
 Matches 93; Conservative 57; Mismatches 152; Indels 3; Gaps 1;

QY 11 RWVGGHPCFIILAIQNHQEDLVAKRMIMAKGADCAKFKSELEFNRKALREPY 70  
 DB 12 RSVGNHEPLIICEIGINBESGLKTFEVDVAVNAAGAVVHQTHIVDESDKQ-- 69  
 QY 71 YTSKSWKTYGHEKHELEFSDQYRBLQRYAEVGIFFTASGDMEMAVEFLHELVNPF 130  
 DB 70 -VTPGADVSIYEIMERCALNEDEIKLRYEYSKMPITSTPFSRAALRLQRMIDIPAY 128  
 QY 131 KVGSGDTNPPYLEKTKAKGKRPVIVSSGQSMQDTKQVQIVKPLNPNCFQCTSAVYL 190  
 DB 129 KVGSGCNVPIPKLVASFQKPIILSTGKNSIESIKSVETIRAGVPYALHCTNIVFT 188

QY 191 QPEDVNLRTVISEYOKLFPDIPIGYSGH-ETGIAISVAALGAKVLERHTLTDKTKWGS 250  
 DB 189 FREDVRLGANDLSEFPDPAITIGLSHTLONTACIAGVALGSLIERHTDMDRPPDI 248  
 QY 251 SASLSEPEGLAEIVRSVRLVERALGSEPTKOLLPCENACNEKLGKSVAVAKVIEGILLTMD 310  
 DB 249 VCSMPDTEKELKQGAHALKARGKDDITLIGKRTKOPAFASVADKIDKGLLSD 308  
 QY 311 MLYTK 315  
 DB 309 NLWVK 313

## RESULT 8

S39722  
 spore coat polysaccharide synthesis protein spse - Bacillus subtilis  
 N:Alternate names: protein ipa-66d  
 C:Species: Bacillus subtilis  
 C:Date: 07-Oct-1994 #sequence\_revision 26-May-1995 #text\_change 20-Jun-2000  
 C:Accession: S39722; C69717  
 R:Giasser, P.; Kunst, F.; Arnold, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, A.; Rapoport, G.; Danchin, A.  
 Mol. Microbiol. 10, 371-384, 1993  
 A>Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region for  
 A:Reference number: S39655; PMID:95020537; PMID:7934828  
 A:Accession: S39722  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-373 <DNA>  
 A:Cross-references: EMBL:X73124; NID:g413923; PIND:CA51623.1; PID:g413951  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993  
 R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brulliet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 A.; Ehrlich, S.D.; Emmerich, P.T.; Entian, K.D.; Errington, J.; Faltret, C.; Ferrati, B.  
 Nature 390, 249-256, 1997  
 A:Authors: Folger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall  
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A.; Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei  
 Y., M.; Ogawa, K.; Ogimura, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A.; Authors: Schleich, S.; Schreier, R.; Scoffone, F.; Sekiguchi, T.; Sekowska, A.; Seron  
 T.; Willems, P.; Wipert, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; PMID:98044033; PMID:9384377  
 A:Accession: C69717  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-373 <DNA>  
 A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIND:CAB15813.1; PID:g2636322  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: spse  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 19.6%; Score 370; DB 2; Length 373;  
 Best Local Similarity 30.1%; Pred. No. 2,7e-21;  
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;  
 QY 13 VGGQHPGFIATIGONHGDIVAKEMIRMAKCGADCAFPKSELEFKENKALERTPYT 72  
 DB 11 VGGDAEPFIIAEGINNDKLDQAFALIDAAAGADAVAFQX----FQADRYGKDPGL 66  
 QY 73 SKRSWGTGYZHK--RHLESHDQYRELQRYAEVGIFFPASGMDMAVEFLHNAVPPF 130  
 DB 67 YKTAAGDVASIFSLVQSWMPAMWILPLLDYCREKOVITFLSTVCDGASDLQSTSPSAP 126  
 QY 131 KVSQSDTNPPYLEKTKKGRPMVVISGQMSMDTKQVQIVPL-NPNEFCLOCTSAAP 189  
 DB 127 KIASYEINHPLKLYVALRLRPFILFAGAEISDVHEARTTIAEENQAIWECAKXP 186

QY 190 LQEDVNLRTVISEYOKLFPDIPIGYSGH-ETGIAISVAALGAKVLERHTLTDKTKWGS 248  
 DB 187 APPEXSNLSIVPMLAAAPFPAVIGSFSDSHSEPTAPCAAVRLGAKLIEKHTIDKRLPGA 246  
 QY 249 DHSASLEPEGLAEIVRSV-----LVERALGSEPTKOLLPCENACNEKLGKSV 295  
 DB 247 DHSFALNDELKKEWDGIRKTEAEIKOGITKPVSEKLGSSYKTTTAEIRFAVRGI 306  
 QY 296 VAAKXIPEDTILTMQMLTV-KVGE-PRAYPEPEDIFFNLV--GKYLTVVEEDDTIM 346  
 DB 307 FTTPAQKGEAFSEDNIAVLAPGKPPQGLHR-FFELLTSVRAVRDIPADTGIV 360

## RESULT 9

DB67604  
 neub protein, probable [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C:Accession: DB67604  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolc  
 n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.H  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A>Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: A87249; PMID:21173698; PMID:11259647  
 A:Accession: DB67604  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-356 <STO>  
 A:Cross-references: GB:ABD05673; NID:g1342484; PIND:AAK24832.1; GSPDB:GN00148  
 C:Genetics:  
 A:Gene: CC2868  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 Query Match 19.0%; Score 358; DB 2; Length 356;  
 Best Local Similarity 30.1%; Pred. No. 2.3e-20;  
 Matches 102; Conservative 55; Mismatches 170; Indels 12; Gaps 9;  
 QY 1 MFLSELPGRKVGWGPCFIITAEIGONHGDIVAKEMIRMAKCGADCAFPKSELEF 60  
 DB 8 IFFSIEIA-GRTIGADSHSPYICELSGNGLERCLAMVAAADTGCDALIKOTYTD- 65  
 QY 61 KENRALRSPYTKS--W-GKTYGHNKHEPSHDQYRELQRYAEVGIFFPASGMDM 117  
 DB 66 -TITLDVRPEKTHIGIMDRTLYELYEAKHTPEEMAAIFERRARGGVITFSSPPDET 124  
 QY 118 AVEFLHLEINVPPEKVGSDTNPNPYLEKTKKGRPMVVISGQMSMDTKQVQIVKPLN- 176  
 DB 125 AVDLIDLSGAPAFKIASFEAVDLPIKRYAAKXGLIISTGMAMLTMOALDTLSAGA 184  
 QY 177 EMFGLGCTSAAYLPEDVNLRTVISEYOKLFPDIPIGYSGHETGIAISVAALGAKYLE 236  
 DB 186 FGVLLALCVSYIPAPPAANVTVDMAARF-GCIGLSDRHRCRAVAANSLGACAVE 243  
 QY 237 RHITLTKTKWGS DHSASLEPEGLAEIVRSVRLVERALGSEPTKOLLPCENACNEKLGKSV 296  
 DB 244 KHTLPADGDPDAFSLPEAFKALVDOTKRAMALRAHDVYGSB-ATSLFRRSILY 302  
 QY 297 AKVXIPEDTILTMQMLTV-KVGE-PRAYPEPEDIFFNLV--GKYLTVVEEDDTIM 346  
 DB 303 VTADVKAEPYLRANVRSVRG-NGLPPADLDKVLAK 339

## RESULT 10

B81275  
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1317 [imported] - Campylobacter jejuni  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: B81275  
 R:Parthill, J.; Wren, B.W.; Mungall, K.; Kestley, J.M.; Churcher, C.; Basham, D.; Chilli  
 C.W.; Quill, M.; Rajadream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barre  
 Nature 403, 665-668, 2000

A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf  
 A:Reference number: A61250; MUID:20150912; PMID:10688204  
 A:Accession: B81275  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-343 <PAR>  
 A:Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:CA873744.1; PID:9696878  
 A:Experimental source: serotype O2, strain NCTC 11168  
 A:Genetics:  
 C:Gene: neuB3; Cj11317  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.9%; Score 356.5; DB 2; Length 343;  
 Best Local Similarity 29.9%; Pred. No. 2.8e-20;  
 Matches 104; Conservative 65; Mismatches 138; Indels 41; Gaps 10;

Db 20 FIIAEGQNGQDLDVAKRMIRAKCGACAKAFQ-----KSELERK---FKR 64  
 13 FIIAEISANMAGSLFVALKSIKAKKAGADAIKIQTYPDLSLTNSDKEDFIIKGLMDK 72  
 Qy 65 KALERPTS---HSMGKTGGEKRLHESHDQYRELQRYAEVGIFFTASGMDENAVFP 121  
 Db 73 RKIYEYSKATYEM-----HSQIFET---NOMEGILCFSSPPAKEDVIF 115  
 Qy 122 LHEINVPFFVYSGDNNPPEYLKTAKEPMWISSGMSMDTMKQVQIVK-PLNPNC 180  
 Db 116 LKRFDPALYKIAFENANDENFRLIKKPKPTIVSGIATEELFICELFEEKNPDIY 175  
 Qy 181 FLQCTGAVLPQPDVNLRVISEQKLPDPIGSGHEGTGIAISVAVALGAKVLEHRT 240  
 Db 176 FLKCTGYTPAIDMMLKGIYSKKEF-NVEGLSHSREFLAPVAVALGARVIRKEM 234  
 Qy 241 LDKTWKSGDHSALREPELAEIVRSYLRALGSPTKOLLPCENAKNEKSGVAAKYK 300  
 Db 235 LDRSISEBDSKFLDPDEFKAMVDARQASALGD-GKLDLDEKVLKRVFARSLYASKD 293  
 Qy 301 IPEGTILTMMLVYKGEPR-AAPPEDIFNLVYKKNLVYVEEDDTIME 347  
 Db 294 IKXGEMFSES--NKSVRPSFGIHPKFXQELLGNKASKDKIDFDALQ 339

RESULT 11  
 H81318  
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1141 (imported) - *Campylobacter jejuni*  
 C:Species: *Campylobacter jejuni*  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: H81318  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett  
 Nature 403, 665-668, 2000  
 A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf  
 A:Reference number: A61250; MUID:20150912; PMID:10688204  
 A:Accession: B81318  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-343 <PAR>  
 A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CA873336.1; PID:9696857  
 A:Experimental source: serotype O2, strain NCTC 11168  
 A:Genetics:  
 C:Gene: neuB1; Cj1141  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.8%; Score 354; DB 2; Length 343;  
 Best Local Similarity 30.5%; Pred. No. 4.4e-20;  
 Matches 96; Conservative 57; Mismatches 156; Indels 6; Gaps 3;

Db 21 IIAEIGNHOGDLDVAKRMIRAKCGACAKAFQSKSELEFKFRKALE-REYTSKHSWGK 79  
 17 IIEIGINNGSLIAIKLMDAARAGAKIINKOTHIIVEDMSQEAQKAVIPGANIS--- 73  
 Qy 80 TYGHHKHLFESHQYRELQRYAEVGIFFTASGMDENAVFELHEINVPFFKVGSGDTNN 139

Db 74 -IYEIMEQCALNYKDELALKEVEKQGLVYLSTPSSRAANRLBDMGSAVYKISGGECCNN 132  
 Qy 140 FPYIAETAKKQRPWYISSGMSMDTMKQVQIVKPLNPNFCLOCTSAVPLQPEVNLRV 199  
 Db 133 YPLIHIAQFKKPMIISTGMSISEIKFTVILIRYEIPFVILHTNLNLYPTSHVRLQA 192  
 Qy 200 ISEYOKLPDPIPIGSGHEGTGIAISVAVALGAKVLEHRTLDKTMKSGDHSALREPEL 259  
 Db 193 KMEIYKFR-NCLYGSDDHTNNALCIGALALGASVLEHRTLDKTMKSGDHSALREPEL 251  
 Qy 260 AEIVSVYLRALGSPFKOLLPCENAKNEKSGVAAKVIPEGTILTMMLVYKGEPR 319  
 Db 252 KDLINQTEMLLRDNNKNPLKEQVTTIDPAPASVSIKDIKSGEILSMQINIVKRSK 311  
 Qy 320 KAYPPEDIFNLVGRK 334  
 Db 312 GGISAKDPEALIGKR 326

RESULT 12  
 B64542  
 spore coat polysaccharide biosynthesis protein E - *Helicobacter pylori* (strain 26995)  
 C:Species: *Helicobacter pylori*  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 20-Jun-2000  
 C:Accession: B64542  
 R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khair, H.G.; Glodek, A.; McKenn  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L  
 Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, M.S.; Borodovsky, M.; Karpi, P.D.; Smith, H.O.; Fraser, C  
 A>Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: B64542  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-340 <TOM>  
 A:Cross-references: GB:AE000538; GB:AE000511; NID:92313263; PIDN:AAD07248.1; PID:9231326  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 16.1%; Score 303; DB 2; Length 340;  
 Best Local Similarity 27.7%; Pred. No. 4.3e-16;  
 Matches 91; Conservative 66; Mismatches 149; Indels 22; Gaps 10;

Db 18 PGTIAEGQNGQDLDVAKRMIRAKCGACAKAFQ-----SELEFKFRKALERPT 72  
 Db 4 PKIYAEISANMAGSLFVALKSIKAKKAGADAIKIQTYPDLSLTNSDKEDFIIKGLMDK 63  
 Qy 73 SKHSWGKTGGEKRLHESHDQYRELQRYAEVGIFFTASGMDENAVFELHEINVPFFK 132  
 Db 64 DRENIEYELQKASTPLEW-HAELEFELAKTL-DIGIF--SSPSSQALTELESLANCPMYKI 119  
 Qy 133 GSGDTNNPPEYLKTAKEPMWISSGMSMDTMKQVQIVKLEPNF---CFLOCTSAVPL 190  
 Db 120 ASPEIYDLDLEKARATQKPIITLSSGIALHTTLQDAISCRVY-NFDITLAKQSAVYS 178  
 Qy 191 QPEDVNLRVISEYOKLPDPIPIGSGHEGTGIAISVAVALGAKVLEHRTLDKTMKSGD 250  
 Db 179 KIEDANLISMYGLSEIF-GVRFGLSDHTIGSLPIATLTIGASMEIKHFIINKSLQTPDS 237  
 Qy 251 SASLEPGLAEIVRSYLRALGSPTKOLLPCENAKNEKSGVAAKVIPEGTILT-- 308  
 Db 238 ASMSQNGKSVNVAIKQSVLALGSEEPINPKTEKRFPARSLFVIRKIDQIGKALTEN 297  
 Qy 309 -MDMLTVKVG-EPKAYPPEDIFNLVGRK 334  
 Db 298 NIKALRPNGIHPKFRK-----EILGQK 320

RESULT 13  
 B71965  
 stalic acid synthase - *Helicobacter pylori* (strain J99)  
 C:Species: *Helicobacter pylori*

A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Jun-2000  
 C:Accession: B71965  
 R:Alm, R.A.; Liang, L.S.; L., M., D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
 ; Ives, C.; Gibson, R.; Merberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Voyts, G.F.;  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A:Reference number: A71800; PMID:99120557; PMID:9923682  
 A:Accession: B71965  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-340 <ARN>  
 A:Cross-references: GB:AE001455; GB:AE001459; NID:G4154678; PIDN:AAD05747.1; PID:G415468  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: neuB  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 15.2%; Score 287; DB 2; Length 340;  
 Best Local Similarity 27.8%; Pred. No. 7, 8e-15;  
 Matches 92; Conservative 68; Mismatches 137; Indels 34; Gaps 14;

QY 21 TTAETGQNHQGDIDVAKRMTAKKCEGACCAKFKQ-----SELEFRNKALEPYTSH 75  
 DB 7 IVAELSANRQDNLAKESLHAIKESGADFKLQYTPSCMTLDSK-----EDPPIIQG 60  
 QY 76 S-WGR--TYGHRK--HLEFSDQYRELQRYAEVIGIFFTASGMDMAVEFLHEINVPF 129  
 DB 61 TMDKQENLYGVOKASTPLEW-HAEFLAKKL-DIGIF--SSPSSKSLLESLDCPM 116  
 QY 130 FVVGSGDTNNPYLEETAKKRPVIVSGQMSQMTWKQYQVKNPNF--CFLOCTSA 187  
 DB 117 YXIASFEIYDLDLIEKARTQKRIILSSGIATAELODAISCRGN-NFDITLLKCYSA 175  
 QY 188 YLPQEDVNLRVISEYKQFPDIPICYSHEGTGIAISVAALGAKVLEHRTLLDKTWG 247  
 DB 176 YSKIEDAHLKSVKLGERT-GYKFLSHHTIGSLPILATLIGASMIKHEILKNSQ 234  
 QY 248 SDHSASLEPGEIAELVRSVLEFALGSPTKQLLPCEMACNEKLGKSVAKYKIPETGL 307  
 DB 235 PSASFMDTFNGFKMGAIKQSVLALGEEBPKINPTELERRFARSLFVIDIQGKAL 294  
 QY 308 TMD--MLTVKVG-EPKAVPEIDIFNLGKK 334  
 DB 295 TSDNIRALRPNDLGFPRYK-----EILGQK 320

RESULT 14  
 H71307  
 probable spore coat polysaccharide biosynthesis protein (spas) - syphilis spirochete  
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 19-Apr-2002  
 C:Accession: H71307  
 R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 ; rson, J.; Khakhria, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD  
 ; they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A:Reference number: A71250; PMID:9832770; PMID:9665876  
 A:Accession: H71307  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-378 <COL>  
 A:Cross-references: GB:AB001232; GB:AB000520; NID:G3322856; PIDN:AAC65539.1; PID:G332285  
 A:Experimental source: strain Nichols  
 C:Genetics:  
 A:Gene: TPO562  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 12.0%; Score 226; DB 2; Length 378;  
 Best Local Similarity 25.1%; Pred. No. 5, 4e-10;  
 Matches 85; Conservative 47; Mismatches 166; Indels 40; Gaps 9;

QY 8 CQBWGGQHPFCFIATIGQNHQGDIDVAKRMTAKKCEGACCAKFKQSELEFRNKALE 67  
 DB 4 CGGCFRPPADILITIAIGSAHNSFRRABALIDAAADAAAKKQ-----LIVASEI 57  
 QY 68 ERYPTSK--HSGWKTYGHRKHLFESHDQYRELQRYAEVIGIFFTASGMDMAVEFLHE 124  
 DB 58 LHPITGVAVRPLSGAVSVYQREELVPLSFYAOQFNARSRGMVIGISPPGSRATATA 117  
 QY 125 LNVPPFVVGSGDTNNPYLEET--AKKGRPMVIVSGM-----QSMQMTWKQYQVKNPNF 179  
 DB 118 LKPFILKVASSEL-NYPTLISTATAAELPLISGCVGLKEIBGALAQCQYTKQ--GSSH 175  
 QY 180 CFLOCTSAVPLQPEDVNLRVISEYKQFPDIPICYSHEGTG-IAISVAALGAKVLEH 238  
 DB 176 ALHLCITAVPAPREYVIMIALPALATIF-NINVSQSHSVDPILVPLIARHAGACTIVEK 234  
 QY 239 ITLQKTMKSGDHSASLEPGEIAELVRSV-----RLVERALGSP 276  
 DB 235 ICLSRIDAGLDDSDIALDPADERTMTALNSCARSPQIISFLHERGAVPVAIVGSG 294  
 QY 277 TKQILPCEMACNEKLGKSVAKYKIPETGLITMDMLTV 314  
 DB 295 EKVLAPSERAHYOKSNRSILHYLAHVPRTVLOKENLVI 332

RESULT 15  
 B80173  
 hypothetical protein aroG [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 18-Nov-2002  
 C:Accession: B80173  
 R:She, Q.; Singh, R.K.; Contaloni, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Char  
 ; Jorg, I.; Jeffries, A.C.; Kozera, C.U.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
 ; arett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, U.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: B80173  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-331 <KUR>  
 A:Cross-references: GB:AE006641; NID:G13813447; PIDN:AAK40641.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: aroG  
 C:Superfamily: phospho-2-dehydro-3-deoxyoctonate aldolase

Query Match 6.5%; Score 122.5; DB 2; Length 331;  
 Best Local Similarity 23.9%; Pred. No. 0.059;  
 Matches 66; Conservative 46; Mismatches 103; Indels 61; Gaps 15;

QY 18 PCFIATIGQNHQGDIDVAKRMTAKKCEGACCAKFKQSELEFRNKALEPYTSHKSW 77  
 DB 95 PCAV-----HNEBQVLTAKAV-----KRAQASL-----RGAVKPRTPSPSF 133  
 QY 78 KTYGHRKHLFESHDQYRELQRYAEVIGIFFTASGMDMAVEFLHEINVPFKGSDT 137  
 DB 134 -GGLGE-----EGVKILRRVQDEVGILVTEIMDTRDSNIFSQY-VMIOIGASNA 182  
 QY 138 NNPFLBETAKKRPMTIVSGQMSQMTWKQYQVKNPNF-----NPNFC-----FLQCTSA 186  
 DB 183 QNFSLIKREVGLKGVLEKRGNG-NYBEMLOAAEYILLBGNNGTVICGEGIRTFEEST 240  
 QY 187 AVPLQPEDVNLRVISEYKQFPDIPIT-GYSGHETG-----IAISVAALGAKVLEH 238  
 DB 241 RFTL---DIGGVAA---KLMTHTLPICADPSHPGKREIVSHLAAVAAAGADMLLEVH 294  
 QY 239 ITLQKTMKSGDHSASLEPGEIAELVRSVRLVERALG 274  
 DB 295 PHEKAL--SDSEQQLTPESFEVLMNRITLAKALG 328

Search completed: May 6, 2004, 09:09:45  
 Job time: 16.8862 secs

Thu May 13 11:53:10 2004

us-09-930-440b-6.rpt

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using SW model

Run on: May 6, 2004, 08:56:04 / Search time 9.44047 Seconds  
(without alignments)  
1980.112 Million cell updates/sec

Title: US-09-930-440B-6

Perfect score: 1887  
Sequence: 1 MPELELCPRGWG3QHPCF.....EEDDTIMEELVDNKGKIKS 359

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883	99.8	359	1	SIMS_HUMAN
2	619	32.8	337	1	YAES_METVA
3	370	19.6	373	1	SPSE_BACSU
4	113.5	6.0	1248	1	SMIB_MOUSE
5	110	5.8	134	1	ANP3_RHIDE
6	108.5	5.7	633	1	PARF_WYCGE
7	108	5.7	63	1	ANP2_AUSBR
8	106.5	5.6	338	1	AROF_THEMA
9	106.5	5.6	478	1	GATB_AQUAS
10	102.5	5.4	1140	1	YD76_MYCPN
11	102	5.4	63	1	ANP1_AUSBR
12	101	5.4	1107	1	MYLE_RAT
13	100.5	5.3	469	1	BGL1_BACST
14	100	5.3	1109	1	WYLE_HOMAN
15	98.5	5.2	371	1	GPR_BACME
16	97.5	5.2	642	1	DNAR_XANCP
17	97	5.1	378	1	ARGE_VIBPA
18	96.5	5.1	641	1	DNAR_XANAC
19	96.5	5.1	1437	1	DPO3_LACPL
20	96	5.1	880	1	PODK_RICPR
21	95.5	5.1	513	1	CHAB_MARPO
22	95.5	5.1	2869	1	RBP1_PLAAB
23	94	5.0	64	1	ANP2_RHIDE
24	93.5	5.0	512	1	GPW1_CLOPE
25	93	4.9	64	1	ANP1_RHIDE
26	93	4.9	542	1	CYMO_AICSP
27	92.5	4.9	1633	1	YP74_CABEL
28	92	4.9	346	1	BPH1_BURCE
29	92	4.9	638	1	DNAR_XYLF
30	92	4.9	1656	1	ATG8_YEAST
31	91.5	4.8	458	1	EF12_RHIRA
32	91.5	4.8	458	1	EF12_RHIRA
33	91	4.8	91	1	ANP3_MACAM

34	91	4.8	197	1	CLPP_BUCAP	Q8K930 buchnera ap
35	90	4.8	538	1	CP18_DROME	Q85078 dirosophila
36	89.5	4.7	381	1	NCAP_CVCAI	P36298 canine ente
37	89.5	4.7	397	1	LE11_METKA	Q81w28 methanopyru
38	89.5	4.7	593	1	MDLB_BOOLI	P75706 escherichia
39	89.5	4.7	638	1	DNAR_XYLF	Q87ba8 xyliella fas
40	89	4.7	838	1	EPF1_XANCP	P45597 x multilipos
41	88.5	4.7	587	1	YVW8_YEAST	P40151 saccharomyc
42	88	4.7	66	1	ANP1_LYCPO	P24028 lycodes pol
43	88	4.7	479	1	SCR8_STIRU	P13522 streptococc
44	88	4.7	508	1	CHLB_SYMP7	Q8gjh0 synechococc
45	88	4.7	855	1	ORP3_MOUSE	Q9dba9 mus musculi

## ALIGNMENTS

RESULT 1	SIMS_HUMAN	STANDARD	PRT	359 AA.
ID	SIMS_HUMAN	Q9NR45; Q9NR49; Q9NR56; Q9NRV4		
AC	Q9NR45; Q9NR49; Q9NR56; Q9NRV4			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	acetylneuraminic acid synthase (N-acetylneuraminic acid synthase) (EC 2.5.1.56) (N-acetylneuraminic acid synthase) (EC 2.5.1.57) (N-acetylneuraminic acid synthase)			
DE	SAS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RC	TISSUE=Liver;			
RX	MEDLINE=20298869; PubMed=10749855;			
RA	Lawrence S.M., Huddleston K.A., Pitts L.R., Nguyen N., Lee Y.C.,			
RA	Vann W.F., Coleman T.A., Betenbaugh M.J.;			
RT	"Cloning and expression of the human N-acetylneuraminic acid phosphate			
RT	synthase gene with 2-keto-3-deoxy-D-glycero-D-galacto-nomonic acid			
RT	biosynthetic ability."			
RL	J. Biol. Chem. 275:17869-17877 (2000).			
RM	[2]			
RP	SEQUENCE FROM N.A.			
RA	Isegaki T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.;			
RA	Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.;			
RA	Nakamura Y., Nagahara K., Masuko Y., Sasaki N.;			
RT	"NEO human cDNA sequencing project."			
RT	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.			
RM	[3]			
RP	SEQUENCE FROM N.A., AND VARIANT ASP-68.			
RC	TISSUE=Liver, and Placenta;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shuman C.M., Schuler G.D.;			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.;			
RA	Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh P.;			
RA	Diachenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.;			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.;			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Mullany S.J.;			
RA	Rata S.S., Loggellano N.A., Peters G.J., Abramson R.D., Gunnarsson P.H.;			
RA	Boesak S.A., McQueen P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.;			
RA	Rickards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W.;			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.;			
RA	Falvey J., Helton B., Kettelman M., Madan A., Rodriguez P., Prange C.;			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.;			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.;			
RA	Butterfield J.S.N., Krzywicki M.I., Skalska U., Smallus D.E.;			
RA	Schneier A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

- FUNCTION: Produces N-acetylneuraminic acid (Neu5Ac) and 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). Can also use N-acetylmannosamine 6-phosphate and mannose 6-phosphate as substrates to generate phosphorylated forms of Neu5Ac and KDN, respectively.

- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine + H<sub>2</sub>O = phosphate + N-acetylneuraminic acid.

- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine 6-phosphate + H<sub>2</sub>O = N-acetylneuraminic acid + 9-phosphate + phosphate.

- TISSUE SPECIFICITY: Ubiquitous.

- SIMILARITY: Contains 1 AFP-like domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF257466; AAF75261.1; -  
 EMBL; AK001659; BAA91818.1; -  
 EMBL; BC000008; AAH00008.1; -  
 EMBL; BC019315; AAH19315.1; -  
 MIM; 605202; -  
 HSP; P19614; NAME.  
 GO; GO:0005737; C:cytoplasm; NAS.  
 GO; GO:0006781; F:N-acetylneuraminic acid synthetase acti. ; NAS.  
 GO; GO:0009103; P:lipopolysaccharide biosynthesis; NAS.  
 InterPro: IPR006014; Antifreeze dom.  
 InterPro: IPR006190; Antifreeze-like.  
 Pfam; PF01354; Antifreeze; 1.  
 Pfam; PF01354; Antifreeze; 1.  
 ProDom; PD003258; Antifreeze1; 1.  
 ProSite; PS00844; AFP\_LIKE; 1.  
 Transfaser; Polymorphism.  
 DOMAIN 294 353 AFP-LIKE.  
 VARIANT 68 68 E -> D.  
 VARIANT 68 68 /FTLD-VAR 013308.  
 CONFLICT 232 232 A -> T (IN REF. 2).  
 CONFLICT 321 321 G -> A (IN REF. 1).  
 SEQUENCE 359 AA; 40307 MW; 2E02D47F4F98592F CRC64;

Query Match 99.8%; Score 1883; DB 1; Length 359;  
 Best local similarity 99.7%; Pred. No. 7, 2e-136;  
 Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELELECPGAWGQHPCTIAGIIGNHGDLDVAKMIMAKECGADCAKPKSELEF 60  
 DB 1 MELELECPGAWGQHPCTIAGIIGNHGDLDVAKMIMAKECGADCAKPKSELEF 60  
 QY 61 KFNRALEPPTSKSKWKTGEGHRLFEFSHDYREIQRVAEEVGIPTASGDENAVE 120  
 DB 61 KFNRALEPPTSKSKWKTGEGHRLFEFSHDYREIQRVAEEVGIPTASGDENAVE 120  
 QY 121 FLHEINVEFFKVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQVQIVPLNPF 180  
 DB 121 FLHEINVEFFKVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQVQIVPLNPF 180  
 QY 181 FLQCSAAPPLOPEDNLEVISYQQLPDPPIGSGHETGIAISVAALAKVLERIT 240  
 DB 181 FLQCSAAPPLOPEDNLEVISYQQLPDPPIGSGHETGIAISVAALAKVLERIT 240  
 QY 241 LDXTKSGDHSASLEPGEIABELVRSVLVERALGSPTRQLPCENACNEKLGKSVANV 300  
 DB 241 LDXTKSGDHSASLEPGEIABELVRSVLVERALGSPTRQLPCENACNEKLGKSVANV 300  
 QY 301 IPEGITLMDLTVVYGEKAYPPEDINLVAKKLVVVEEDDTIMELVNEGKRTKS 359  
 DB 301 IPEGITLMDLTVVYGEKAYPPEDINLVAKKLVVVEEDDTIMELVNEGKRTKS 359

RESULT 2  
 YAE5 METVA STANDARD; PRT; 337 AA.  
 ID YAE5 METVA  
 AC 058465;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein MJ1065.  
 GN MJ1065.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 CC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCST\_TaxID=2190;  
 [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=868087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Klerke A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Nierlich J.M., Glodex A.,  
 RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fuhrman J.L., Nguyen D.,  
 RA Urtreback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Corton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";  
 RL Science 273:1056-1073 (1996).  
 CC -1- SIMILARITY: STRONG, TO B.SUBSTITUS SPSE.  
 CC -1- SIMILARITY: Contains 1 AFP-like domain.  
 CC -1- SIMILARITY: Contains 1 AFP-like domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U67549; AAB99068.1; -  
 PIR; H64432; H64432.  
 DR TIGR; MJ1065; -  
 DR InterPro: IPR006014; Antifreeze dom.  
 DR InterPro: IPR006190; Antifreeze-like.  
 DR InterPro: IPR004144; Neus.  
 DR Pfam; PF01354; Antifreeze; 1.  
 DR Pfam; PF01354; Antifreeze; 1.  
 DR PROSITE; PS00844; AFP\_LIKE; 1.  
 DR Hypothetical protein; Complete proteome.  
 FT DOMAIN 279 337 AFP-LIKE.  
 SQ SEQUENCE 337 AA; 37976 MW; 790CCAPF48C1111B CRC64;

Query Match 32.8%; Score 619; DB 1; Length 337;  
 Best local similarity 40.0%; Pred. No. 8, 1e-40;  
 Matches 136; Conservative 65; Mismatches 115; Indels 24; Gaps 7;

QY 11 RWVGQHPCTIAGIIGNHGDLDVAKMIMAKECGADCAKPKSELEFKNKALERP 70  
 DB 9 RWVGQHPCTIAGIIGNHGDLDVAKMIMAKECGADCAKPKSELEFKNKALERP 70  
 QY 71 YTSKSGSKTGYGHHKHLFEFSHDYREIQRVAEEVGIPTASGDENAVEFLHEINVEFF 130  
 DB 59 FISKYS---EYELFKSLSESEFFHLKRYAKIGIMSTFLDKYVILNKMPAF 115  
 QY 131 KVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQVQIVPLNPF---FLQCS 186  
 DB 116 KIASGDLTFPPLKRYAKIGIMSTFLDKYVILNKMPAF 172  
 QY 187 AYPLOPEDNLEVISYQQLPDPPIGSGHETGIAISVAALAKVLERITLMDTVK 246  
 DB 173 SYPLPYSDVNLAKIKTKSLF-NIPVGISDHTIGILAPVAVSALGADVIERHFTLDKMS 231

QY 247 GSDHSLERGLAEIVRSLVERALGSPKPOLPCMACNEKLGKSVAKYKPIPECTI 306  
 DB 232 GSDHSLADPEEFKVMNRIIVERKLGSGEKIIPSESDVIVEARSLVAKENIKGKY 291  
 QY 307 LTMDELTVKVGEP-KAYPPEDIFNLTGKVLTVVEEDTI 345  
 DB 292 LSVDNISFK-RPRGIEFTKYSLIIMRKIKNDKEEDII 329  
 RESULT 3  
 SPSE\_BACSU STANDARD; PRT; 373 AA.  
 ID SPSE\_BACSU 393625;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Spore coat polysaccharide biosynthesis protein spsB.  
 GN SPSE OR IPA-67D OR BS037870.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_Taxid=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95020537; PubMed=7934828;  
 RA Glaeser P., Kunst F., Arnold M., Condart M.P., Gonzales W.,  
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,  
 RA Pressac E., Santana M., Schneider B., Schweizer J., Vertes A.,  
 RA Rapoport E., Danchin A.;  
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
 kb region from 325 degrees to 353 degrees."  
 RL Mol. Microbiol. 10:371-384 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Broutin L., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denliac F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Euteneier K.D., Errington J., Fabret C., Ferrari B., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleton N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grand G.,  
 RA Gualseppe G., Guy B.J., Hage K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Katamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Maunda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno K., Mostl D., Nakai S., Nohack M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portetelle D., Potworlik S., Prescott A.M.,  
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield P.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Solido B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosiato V., Uchiyama S., Vandenbol M., Vamler F., Vassarotti A.,  
 RA Vassari A., Wambut R., Wedler H., Wedler H., Weitenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zurekstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis".  
 RL Nature 350:249-256 (1997).  
 CC -1- PATHWAY: Spore coat polysaccharide biosynthesis.  
 CC -1- SIMILARITY: Contains 1 AFP-like domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: X73124; CAAS1623.1;  
 DR EMBL: 299123; CAB15013.1;  
 DR PIR: S39722; S39722.  
 DR Subtilin; BG10613; spsB.  
 DR InterPro: IPR006014; Antifreeze dom.  
 DR InterPro: IPR006190; Antifreeze-like.  
 DR InterPro: IPR004144; Neut.  
 DR Pfam: PF01354; Antifreeze; 1.  
 DR Pfam: PF01302; Neut; 1.  
 DR PROSITE: PS50844; AFP-LIKE; 1.  
 KW Complete proteome.  
 FT DOMAIN 305 367 AFP-LIKE.  
 SQ SEQUENCE 373 AA; 40889 MW; 3CFBDFACE0DCE38 CRC64;  
 Query Match 19.6%; Score 370; DB 1; Length 373;  
 Best Local Similarity 30.1%; Freq. No. 7.8e-21;  
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;  
 QY 13 VGGGPPCTIARIIGNHGGDIDVAKMIRMAKCGADCAKFKSLEFFNKALEPYT 72  
 DB 11 VKADAPVITIAAGINHGKLDQALIDAAAEAGADAVKQM-----FQADMVYKDPOL 66  
 QY 73 SGRWNGKTYGERK-RHLEFSDQYRELQRYAEVGIFFTASGMDENAVFHELVPEF 130  
 DB 67 YXTAGKQVSIPISTVQSMEMPALWILPLDYCREKQVIFLSTVCDEGSDLLQSTSPSAF 126  
 QY 131 KVGSDTNNPFLKRTAKGPRMVISGQMSMDTKQYVQIYKPL-NPNPCTLOCTSAVP 189  
 DB 127 KLAETINHLPLKTVARNRPMPTSTNGAETSDVHAKRITRAAGNOAIAMHCYAKP 186  
 QY 190 LQPEDVNLVISEYOKLPPDIPIGYGH-ETGIAISVAVALGAKVLEHITLDTKWSG 248  
 DB 187 APPEYNSLVIMLAAPPEAVTGFSDSHETPAACAVRGLATIEHFTIDKMLPGA 246  
 QY 249 DHSALFPELAEVRSVR-----LVRAAGSPKPOLPCMACNEKLGKSV 295  
 DB 247 DHSFALNPELKEWMDGKRTAEIKQGITKPVSKLSSYKTTTALBGRINRAVKRI 306  
 QY 296 VAKVPIBEGITLTMULTY-KVGE-PKAVPPEDIFNLV-GKVLTVVEEDTI 346  
 DB 307 FTTAPIQKAESEINIAVLRGCKPQGIHPR-FHELTGVRVADIPADTGIV 360  
 RESULT 4  
 SMB\_MOUSE STANDARD; PRT; 1248 AA.  
 ID SMB\_MOUSE 0920F6;  
 AC 0920F6;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Structural maintenance of chromosomes 1-like 2 protein (SMC1beta  
 DE protein).  
 GN SMC12.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, FUNCTION, CHARACTERIZATION, AND  
 RP INTERACTION WITH SMC3.  
 RC TISSUE-Testis;  
 RX MEDLINE=21448993; PubMed=11564881;  
 RA Revenkova E., Bipe M., Heyling C., Jessberger R.;  
 RT "Novel meiosis-specific isoform of mammalian SMC1".  
 RL Mol. Cell. Biol. 21:6984-6998 (2001).  
 CC -1- FUNCTION: Meiosis specific component of cohesin complex. The  
 CC cohesin complex is required for the cohesion of sister chromatids  
 CC after DNA replication. The cohesin complex apparently forms a

large proteinaceous ring within which sister chromatids can be trapped. At anaphase, the complex is cleaved and dissociates from chromatin, allowing sister chromatids to segregate. The meiosis-specific cohesin complex probably replaces mitosis specific cohesin complex when it dissociates from chromatin during prophase I.

-1- SUBUNIT: Forms a heterodimer with SMC3. Component of a meiosis-specific cohesin complex, probably composed of the SMC2L2 and SMC3 heterodimer attached via their hinge domain, RAD21 (or its meiosis-specific related protein REC8), which link them, and STAG3, which interacts with RAD21 or REC8.

-1- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin. In prophase I stage of meiosis, it is found along the axial elements of synaptonemal complexes. In late-pachyene-diplotene, the bulk of protein dissociates from the chromosome arms probably because of phosphorylation by PLX, except at centromeres, where cohesin complexes remain. It however remains chromatin associated at the centromeres up to metaphase II. At anaphase II, it dissociates from centromeres, allowing chromosome segregation.

-1- TISSUE SPECIFICITY: Testis and ovary specific. Not expressed in somatic cells.

-1- DOMAIN: The flexible hinge domain, which separates the large intramolecular coiled coil regions, allows the heterocyclic interaction with the corresponding domain of SMC3, forming a V-shaped heterodimer. The two heads of the heterodimer are then connected by different ends of the cleavable RAD21 or REC8 protein, forming a ring structure (By similarity).

-1- SIMILARITY: Belongs to the SMC family. SMC1 subfamily.

-----

CC This SMS-S-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: AF303827; MAF09333.1; --

DR MGD: MGI:2154049; SMC112.

DR GO: GO:0000775; C:chromosome, pericentric region; IDA.

DR GO: GO:0000795; C:synaptonemal complex; IDA.

DR GO: GO:0005177; F:DNA binding; IDA.

DR GO: GO:0005151; F:protein binding; IPI.

DR GO: GO:0007062; P:sister chromatid cohesion; IDA.

DR InterPro: IPR003439; ABC transporter.

DR InterPro: IPR003405; SMC\_C.

DR InterPro: IPR003395; SMC\_N.

DR Pfam: PF02483; SMC\_C; 1.

DR Pfam: PF02463; SMC\_N; 1.

DR Meiosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil; Nuclear protein.

KM NP BIND 32 39 ATP (POTENTIAL).

FT DOMAIN 163 502 COILED COIL (POTENTIAL).

FT DOMAIN 503 665 FLEXIBLE HINGE.

FT DOMAIN 666 912 COILED COIL (POTENTIAL).

FT DOMAIN 1123 1158 ALA/ASP-RICH (DA-BOX).

FT DOMAIN 1248 144513 MW: 6C31DB46217BC94D CRC64;

SEQUENCE

Query Match 6.0%; Score 113.5; DB 1; Length 1248;

Best local similarity 23.5%; Pred. No. 1.1;

Matches 88; Conservative 56; Mismatches 138; Indels 93; Gaps 19;

QY 31 GDLIVAKKRIIRNAKCGACACAFQKSELEFKRNR--ALERTYKSH--NKTGYGHR 86

DB 166 GEYBAKKKLQVRAE-----DAQFHVKKVNAER--KAKKIEKEEAHQ 210

QY 87 H-----LEPSHDQYRELRQYAEVGIPTASGDMVAVEFLHEINVPFVAGSDT--NMP 141

DB 211 NLLEELIKIKIKIQLMFLQVYNEKIVLVNMLEQ-----DGNLSVAK-----DTLSHR 260

QY 142 YLETKAKGPPVYISSGMSMDTMRKQYQIVQVLANFGLQCTSAVPLQPEDVNLKVS 201

DB 261 NIFKAKKQVGMTRQLOQTAKELKSEVAILNQRQYTKAVENTSHLKLKLSKLLT 320

QY 202 EYQKLPDIPIGYSGHETGIAISVAVALKVLERHITLDTKTKGSDHSAS---LEPGE 258

DB 321 DNEK-----QSKQSDGIRALVAEL-----DDRWKSPFKQMEKILQKGR 363

QY 259 LAELVRS-----VAVLRALGSPFKQLLPC---EMAGNEKIG-----KS 294

DB 364 DIELNSQLRYKLKQVRAKRGVIMTQQLDKQWBEKAKERLAFKRRHGTQGNLKQ 423

QY 295 VAVRV-----KIPGTLTNDMLTVKVGEPKAPPEDFINVGKVLVYWEEDTITM 346

DB 424 IKQDIEHKRIEKLSEYVTCMDLEDKQGBALKE--IENYKSESVN--EELSILR 481

QY 347 EEL-----VDNH--GRX 356

DB 482 NELQNAQIDNHGKGR 496

RESULT 5

ANP3\_RHIDE STANDARD; PRT; 134 AA.

AC P35753;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Antifreeze peptide RD3.

OS Rhigophila dearborni (Antarctic eel pout) (Lyocodichthys dearborni).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthopterygii; Acanthopterygii; Percomorphi; Perciformes; Zoarcoidei; OC Zoarcoidei; Lyopoda; Acanthopterygii.

OX NCBI\_TaxID=8201;

OX [1]

RP SEQUENCE.

RP MEDLINE=95210308; PubMed=7696304;

RA Wang X., Devries A.L., Cheng C.-H.C.,

RT "Antifreeze peptide heterogeneity in an antarctic eel pout includes an unusually large major variant comprised of two 7 kDa type III AFPs linked in tandem."

RL Biochim. Biophys. Acta 1247:163-172(1995).

RN [2]

RP STRUCTURE BY NMR OF 1-73.

RX MEDLINE=99355369; PubMed=10423534;

RA Mitra K., Ouyiya S., Hoshino T., Nemoto N., Oodaira M., Nitta K., Tsuda S.,

RT "Determination of the solution structure of the N-domain plus linker of antarctic eel pout antifreeze protein RD3."

RL J. Biochem. 126:387-394(1999).

CC -1- FUNCTION: Antifreeze proteins lower the blood freezing point by absorbing ice and inhibiting its growth.

CC -1- SIMILARITY: Belongs to the type-III AFP family.

CC -1- SIMILARITY: Contains 2 AFP-like domains.

CC PIR: S53514; S53514.

DR PDB: 3NLA; 23-FEB-99.

DR PDB: 3RDN; 23-FEB-99.

DR PDB: 1C89; 28-FEB-01.

DR PDB: 1C8A; 28-FEB-01.

DR InterPro: IPR006014; Antifreeze dom.

DR InterPro: IPR006190; Antifreeze-like.

DR InterPro: IPR006013; AntifreezeIII.

DR Pfam: PF01354; Antifreeze; 2.

DR PRINTS: PR00357; ANTIFREEZEIII.

DR ProDom: PD003258; AntifreezeIII; 2.

DR PROSITE: PS00844; AFP\_LIKE; 2.

KM Antifreeze protein; Multigene family; Repeat; 3D-structure.

KM DOMAIN 4 63 AFP-LIKE 1.

FT DOMAIN 65 70 LINKER.

FT DOMAIN 74 133 AFP-LIKE 2.

FT SITE 9 9 IMPORTANT FOR ICE-BINDING (BY SIMILARITY).

FT SITE 14 14 IMPORTANT FOR ICE-BINDING (BY SIMILARITY).

FT SITE 18 18 IMPORTANT FOR ICE-BINDING

FT SITE 44 44 (BY SIMILARITY).  
 FT SITE 79 79 (BY SIMILARITY).  
 FT SITE 84 84 (BY SIMILARITY).  
 FT SITE 88 88 (BY SIMILARITY).  
 FT SITE 114 114 (BY SIMILARITY).  
 FT STRAND 4 7  
 FT STRAND 11 11  
 FT STRAND 19 21  
 FT STRAND 22 25  
 FT STRAND 33 33  
 FT STRAND 37 40  
 FT HELIX 41 42  
 FT TURN 45 45  
 FT STRAND 49 49  
 FT TURN 51 52  
 FT STRAND 53 54  
 FT TURN 57 59  
 FT STRAND 60 60  
 SQ SEQUENCE 134 AA; 14480 MW; 77P208BF3E2CA54 CRC64;  
 Query Match 5.8%; Score 110; DB 1; Length 134;  
 Best Local Similarity 37.4%; Pred. No. 0.13; 37; Indels 8; Gaps 2;  
 Matches 34; Conservative 12; Mismatches 37; Indels 8; Gaps 2;  
 QY 269 VERALGSPKQLPCEVACNEKLG-----KSYVAKKIEGGLTMDLTVKGEPRKAY 322  
 DB 45 VNRAY-PLGTLMPDKVKNYEDGTTSPKSVYANGLIINATLTVKAKBVSPPKGI 102  
 QY 323 PPDIFNLVKKYLVYVEDDTIMEELVDNH 353  
 DB 103 PSEISKLWGVNRVAVYLDQTLMPDKVKNY 133  
 RESULT 6  
 PARE\_MYCGE STANDARD; PRT; 633 AA.  
 AC P47445;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Topoisomerase IV subunit B (EC 5.99.1.-).  
 GN PARE OR MG203.  
 OS Mycoplasma genitalium.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 CX NCBI\_TaxId=2097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE=56026346; PubMed=7569993;  
 RA Frazer C.K., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,  
 RA Nguyen D.T., Uterback T.R., Sauder D.W., Phillips C.A., Merrick J.M.,  
 RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lincer T.S.,  
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.:  
 RT "The minimal gene complement of Mycoplasma genitalium.",  
 RL Science 270:397-403(1995).  
 RP SEQUENCE OF 406-633 FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RA Bailey C.C., Younkins R., Huang W.M., Bort K.F.;  
 RL submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Topoisomerase IV is essential for chromosome  
 CC segregation. It has relaxation of supercoiled DNA activity.  
 CC Performs the decatenation events required during the replication  
 CC of a circular DNA molecule (by similarity).  
 CC -1- SUBUNIT: Composed of two subunits: parC and parE.

CC -1- SIMILARITY: Belongs to the type II topoisomerase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation on  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U39700; AAC71421.1;  
 DR EMBL: U25549; AAC43990.1;  
 DR PIR: D64222; D64222.  
 DR HSSP: P06982; LAJ6.  
 DR TIGR: MG203;  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR002288; DNA\_gyraseB\_C.  
 DR InterPro: IPR001241; DNA\_topoisom.  
 DR InterPro: IPR005740; TopoIV\_B\_Gpos.  
 DR InterPro: IPR006171; Toprim\_dom.  
 DR Pfam: PR00204; DNA\_gyraseB\_1.  
 DR Pfam: PR00986; DNA\_gyraseB\_C\_1.  
 DR Pfam: PR02518; HATPase\_C\_1.  
 DR Pfam: PF01751; Toprim\_1.  
 DR PRINTS: PR00418; TP12FAMILY.  
 DR PRODOM: PD149633; HATPase\_C\_1.  
 DR SMART: SM00433; TOP2C\_1.  
 DR SMART: SM00433; TOP2C\_1.  
 DR TIGRPFAM: TIGR01058; PARE\_Gpos\_1.  
 DR PROSITE: PS00177; TOPOISOMERASE\_II; 1.  
 DR Isomerase; Topoisomerase; ATP-binding; Complete proteome.  
 KW Isomerase; Topoisomerase; ATP-binding; Complete proteome.  
 SQ SEQUENCE 633 AA; 71569 MW; 249B854169A8CB6 CRC64;  
 Query Match 5.7%; Score 108.5; DB 1; Length 633;  
 Best Local Similarity 18.1%; Pred. No. 1.1;  
 Matches 77; Conservative 63; Mismatches 139; Indels 147; Gaps 18;

QY 26 GONHGDLDVANRMTMAKEG-----ADCAFYQS-----ELEPK 61  
 DB 137 GOIYGVYVAGGKIIQAKKIDGTTSGTTVSFPHADPKVFKIAQPSNIIKRLKELSTL 196  
 QY 62 FNRKLERIYTSKSHMGKTGEGKHLBESHQYRELQYASEVGIFPTASGDMANAVEF 121  
 DB 197 FAKLKL-----TFTDQK-----TKTIYFFETSGI-----VGF 224  
 QY 122 LRLNVPFVKVSGGTNNPYLEKTAKGR-----PWYSSGQSGDMTKOYQIVKEL 175  
 DB 225 LDEINNTVETLG-----QKTLIKGKDGIEVEVFORNSDQETILSPANSIKTF 274  
 QY 176 -----NPNRCFL-----QCTSAVPLQPEVDNLRVISEYKL-----PFDIPIGYSG 216  
 DB 275 EGSGHENGFCLASDVINSYCKRNLLKEDKNFQISEIRQGLNALIKRNLEKRIAPFG 334  
 QY 217 HETGIAISVAVALGAKVLERH-----ITLDKTYKSGDSHSLPGLAEIV 263  
 DB 335 QIKSKLFSKEVNVAVYELVQHYFQFLERNNNDAKLIIDKL-----LNARKIKEOI 385  
 QY 264 KSVRLVERALGSPTKO-----LLPCE-----MACNEKLGKSVAKVKI 301  
 DB 386 KQQRLEKSLSPQREKILFGKLAPQYTKSEKELFIVEGSGAGSTAVKGRIRIQALI 445  
 QY 302 P-EGTLTMDLTVKGEPRKAYPEDIFNL-----VGKVLVYVEDDTIMEELV-----DN 352  
 DB 446 PLRGKLVANKEKNNK-----KEALTNEILLILICIGILTNFINIKLKGKIIIMTDAON 502  
 QY 353 ECKKIK 358  
 DB 503 DGAHIQ 508  
 RESULT 7  
 ANP2\_AUSBR STANDARD; PRT; 63 AA.  
 ID ANP2\_AUSBR

RA MEDLINE=99287316; PubMed=10360571; Gwin M.L., Dodson R.J.,  
RA Nelson K.E., Clayton R.A., Gill S.R.,  
RA Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Uetrecht T.R., Malek J.A., Linber K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,  
RA "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of *Thermotoga maritima*.";  
RL Nature 399:323-329(1999).

RA CHARACTERIZATION.  
RP STRAIN=MSB / DSM 3109 / ATCC 45689;  
RC MEDLINE=22753848; PubMed=12743122;  
RA Wu J., Howe D.L., Woodard R.W.;  
RT "Thermotoga maritima 3-deoxy-D-arabino-heptulosonate 7-phosphate  
RT (DHAP) synthase: the ancestral eubacterial DHAP synthase?";  
RL Biol. Chem. 278:27525-27531(2003).

CC -1- FUNCTION: Catalyzes the condensation of phosphoenolpyruvate (PEP)  
CC and D-erythrose-4-phosphate (E4P) giving rise to 3-deoxy-D-  
CC arabino-heptulosonate-7-phosphate (DAH).  
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + D-erythrose 4-phosphate  
CC + H(2O) = 3-deoxy-D-erythro-hep-2-ulosonate 7-phosphate +  
CC phosphate.  
CC -1- COFACTOR: Requires divalent ions.  
CC -1- ENZYME REGULATION: Inhibited by L-phenylalanine and L-tyrosine.  
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
CC first step.  
CC -1- SUBUNIT: Homotetramer.  
CC -1- MISCELLANEOUS: Optimal pH is 6.3 at 60 degrees Celsius. Maximal  
CC activity is at 90 degrees Celsius. Extremely thermostable.  
CC -1- SIMILARITY: Belongs to class-I DAHP synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/  
CC or\\_send\\_an\\_email\\_to\\_license@isb-sib.ch](http://www.isb.ch/announce/or_send_an_email_to_license@isb-sib.ch)).

CC EMBL; AE001715; AAC35429.1; -  
CC PIR; E72388; E72388.  
CC TIGR; TM0343; -  
CC InterPro; IPR006218; DAHP1/KDSA.  
CC InterPro; IPR006268; DAHP\_syn\_2.  
CC Pfam; PF00793; DAHP\_syn\_1; 1.  
CC TIGRFAMs; TIGR01361; DAHP\_synth\_Bsub; 1.  
CC Aromatic amino acid biosynthesis; transferase; Complete proteome.  
SQ SEQUENCE 338 AA; 37378 MW; B9634B6704D3DFAD CRC64;

Query March 5.7%; Score 108; DB 1; Length 338;  
Best local Similarity 22.4%; Pred. No. 0.57;  
Matches 54; Conservative 50; Mismatches 91; Indels 46; Gaps 10;  
Db 118 SELGVTLRGAYPRSPSYF--QDGEKGLYE-----LEADKRGMYVTALG 167  
Cy 56 SELEFKEFKALRRPYTSKSHSGWGTYSGEKRLHFSHDQYRELQRYAVEVGIFPTASGD 115  
Db 116 EMAYEFLHNLNVPFKYSGSDTNPFLETKTAKGRBMVSSGQSGMDTKYQVQIYKPL 175  
Cy 168 EDDLPKFAEY-ADTIQIGARNAPQFLLSLGASYNKFLVLRGF--MTIEEPLSLAEYI 224  
Db 176 --NNNFQLOQ-----TSAYPLQPEDVNLRVISEYKLPEDIDIGTSG- 216  
Cy 225 ANSGNTKILICRGIRTFPKATNTLIDISAVPIIRKSHPLIT-----VDPSSGG 275  
Db 217 -HEGGLISVAAYVGLAK--VLEHHTLDKTKMSDSDASLEPEHLAIVPSVLYERAL 273  
Db 276 RRDIVIPLSRAALVAGHIIIVEHPEPEKAL--SDKQSDLEFELKELVEMKKDAL 333  
Cy 274 G 274

Db 334 G 334

## RESULT 9

DB GATB\_AQUAE STANDARD; PRT; 478 AA.  
AC 066766;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aspartyl-glutamyl-tRNA(Asn/Gln) amidotransferase subunit B  
DE (EC 6.3.5.-) (asp/Glu-ADT subunit B).  
GN GATB OR AQ\_461.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxId=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.D., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358(1998).  
CC -1- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)  
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-  
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both  
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction  
CC takes place in the presence of glutamine and ATP through an  
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (by  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.  
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP  
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.  
CC -1- SIMILARITY: Belongs to the gatB/gatB family. GatB subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AB000691; AAC06727.1; -  
DR PIR; B70342; B70342.  
DR HAMAP; MF 00121; -1.  
DR InterPro; IPR000443; GatB.  
DR InterPro; IPR006107; GatB\_Cent.  
DR InterPro; IPR006075; GatB\_N.  
DR InterPro; IPR003789; GatB\_Yqey.  
DR Pfam; PF01162; GatB\_1.  
DR Pfam; PF02934; GatB\_N\_1.  
DR Pfam; PF02637; GatB\_Yqey\_1.  
DR TIGRfam; TIGR00133; gatB\_1.  
DR PROSITE; PS01234; GATB\_1.  
KM Protein biosynthesis; Ligase; Complete proteome.  
KW SQUENCE 478 AA; 55040 MW; F725AE78944BD79A CRC64;  
SQ  
Query Match 5.4%; Score 106.5; DB 1; Length 478;  
Best local Similarity 19.6%; Pred. No. 1.1; Indels 101; Gaps 14;  
Matches 62; Conservative 54; Mismatches 100; Indels 101; Gaps 14;  
QY 93 DQRELORYAEVGFITAGMDNAVEFHELANVPPFKVSGD-----TNNPPL 143  
DB 171 EKLINIRYA-----SKADWEKQALRCIDINVISIRPKSKEFTREIVEIKVNSFRV 223  
QY 144 EKXAKKGRPVVTSQGMQSDTMKQVQIVKPLNPNCEFLQCTSAVPLQ----- 191

Db 224 QKALE-----YETEROINVEBGEVVOETRTFD-----QTKTTPMRTEKAEDRYRP 274  
QY 192 -PEDVNLRY-----ISEYQKLPFDIP-----IGYSGHETGIAISVAVALGAKYLE 236  
DB 276 DEDLVPLAVKWKEMIEIKKNMPELFDQFERLIKGYGSEYAGILVNHKEVGDFFBEAV 334  
QY 237 REITLTK-----TW-----KG-SHNSASLAEGLABLYASRLVLRALGSPKQLL 281  
DB 335 RHFKEPKGIWMLINDLLGLARDKGISIEESPVKPEHIAELVKLIK----- 380  
QY 282 PCMAKNEKLGKSVYAKVIFPGTILTMCMILVKYGEPPAYPPEDIFNLWKGVIVTVEE 341  
DB 381 --EKVISTKIGVEVKKEM-----VERCK-----TPSQIVEKKGKQITDENQ 420  
QY 342 DDTIMEELVNDGKKIK 358  
DB 421 IELVKKIFEKPKKEVE 437

## RESULT 10

DB MYCPN STANDARD; PRT; 1140 AA.  
AC P75405;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MNP376 (A19\_071140).  
GN MNP376 OR MP460.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxId=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97103885; PubMed=8948633;  
RA Himelreich R., Hildert H., Plegens H., Pirk E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RT pneumoniae";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE  
CC MNP375.  
CC -1- SIMILARITY: IN THE CENTRAL SECTION; STRONG, TO M.PNEUMONIAE  
CC MNP374.  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE  
CC MNP373.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AE000045; AAB96108.1; -  
DR PIR; S73786; S73786.  
DR Hypothetical protein; Transmembrane; Complete proteome.  
KM TRANSMEM 8 28 POTENTIAL.  
FT TRANSMEM 1098 1118 POTENTIAL.  
SQ SEQUENCE 1140 AA; 130383 MW; 8FA9406C57D8886 CRC64;  
Query Match 5.4%; Score 102.5; DB 1; Length 1140;  
Best local Similarity 21.3%; Pred. No. 6.7; Indels 83; Gaps 16;  
Matches 84; Conservative 60; Mismatches 167; Indels 83; Gaps 16;  
QY 22 IABIGNHGDDVAKMIRNA-----KCGADCAKFOK--SELEKFRKALERY 71  
DB 598 VAIIGFQDEGGIRLDSFINKVALTRKSKTKTLANDDASLEKAAEISYKH-----YRQN 653  
QY 72 TSKHSGKTYGGEKHLSESHDQRELORYAEVGFITAGMD--EVAVEFLHELVNF 129  
DB 654 FROASMDKXNSQTKSIWOSTDLNDRERFORDINNYLKVQGISETEIVNVAHVQDA-M 712

```

QY      130 FKXGSDTNNFPLLEKTAKKGRPMWISGMOGMDTMRKVOYQIYVPLNPN-----178
Db      713 LNRKSDDDKRLASVSGTANK-----YGLNRANPNYTGQFYVVVDVYTNANDLGNORANNA 767
QY      179 ---FCLQ-----CTSNAPLOPEDMNLRNISEYQKLFPDIPGYSGHETGLAISVA--- 227
Db      768 KSYFYIEGLDKAOSLYVRE--NRKQLYSLESLAVDSRGLVYKNSXDALIQAKQO 825
QY      228 -----VALGAKYLERHITL-----DKTWKSDSHASLEPEGLAEIVRSVRL--VZ 270
Db      826 NLYLDTHNNMMLAKMLTNAEBITLPPASADNSAKMLSTPAENDBOGLFENVSQSILGYVE 885
QY      271 PALG-----SPTKQILPEMAONEKLGKSVAKYKIPBGTILTMDM-----LT 313
Db      886 RMTGKQLFLKERVSPFKEDNNKMLKLTNFTLDDKKNMLRVDP--SVINQIVESBAKYN 943
QY      314 VKVGEPAKYPREDIPNLYGVKVLVVEEDDTIME 347
Db      944 VLVSEEGKDDPESDKNIP-KITLITTPESSTYIK 976

```

ID	ANPL_AUSBR	STANDARD;	PRT;	63 AA.
AC	PL12100;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DS	Antifreeze peptide AB1.			
DS	Austrolyctichthys brachycephalus (Antarctic eel pout).			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei; Neoteleostei;			
CC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
CC	Acanthomorphi; Acanthopterygii; Percomorphi; Perciformes; Zoarcoidei;			
CC	Zoarcoidei; Austrolyctichthys.			
OX	NCBI_TaxID=8195;			
RN	(1)			
RP	SEQUENCE.			
RX	MEDLINE=89323219; PubMed=2752054;			
RA	Cheng C.-H.C., Devries A.L.;			
RT	"Structures of antifreeze peptides from the antarctic eel pout,			
RL	Austrolyctichthys brachycephalus.":			
CC	Biochim. Biophys. Acta 977:55-64(1988).			
CC	-1- FUNCTION: Antifreeze proteins lower the blood freezing point by			
CC	absorbing ice and inhibiting its growth.			
CC	-1- SIMILARITY: Belongs to the type-III AFP family.			
CC	-1- SIMILARITY: Contains 1 AFP-like domain.			
CC	PIR: S04973; S04973.			
DR	HSSB; P35753; 3RDN.			
DR	InterPro; IPR006014; Antifreeze_dom.			
DR	InterPro; IPR006190; Antifreeze_like.			
DR	InterPro; IPR006013; AntifreezeIII.			
DR	Pfam; PF01354; Antifreeze; 1.			
DR	PRINTS; PRO0357; ANTIFREEZEIII.			
DR	ProDom; PD003258; AntifreezeIII; 1.			
DR	PROSITE; PS00844; AFP_LIKE; 1.			
KW	Antifreeze protein.			
FT	DOMAIN	3	62	AFP-LIKE.
FT	SITE	8		IMPORTANT FOR ICE-BINDING
FT	SITE			(BY SIMILARITY).
FT	SITE	13	13	IMPORTANT FOR ICE-BINDING
FT	SITE			(BY SIMILARITY).
FT	SITE	17	17	IMPORTANT FOR ICE-BINDING
FT	SITE			(BY SIMILARITY).
FT	SITE	43	43	IMPORTANT FOR ICE-BINDING
FT	SITE			(BY SIMILARITY).
SQ	SEQUENCE	63 AA; 6846 MW;	AC84FD14247193B4	CRC64;
Query Match		5.4%;	Score 102;	DB 1; Length 63;
Best Local Similarity		39.3%;	Pred. No. 0.2;	
Matches	24;	Conservative	10;	Mismatches
			27;	Indels
			0;	Gaps

Db 2 KSVVASQILPINALTPANMKAKEVSPKGI PAEMSKITVGMQVNAVRANUDETILAPDMVKI 61

QY 353 H 353

Db 62 Y 62

```

Query Match          5.4%; Score 101; DB 1; Length 1107;

ID MYIE RAT STANDARD; PRT; 1107 AA.
MYIE RAT
AC 06356;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin Ie (Myosin heavy chain myr 3).
GN MYOIE OR MYR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OK NCBI_TaxID=10116;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=95247829; PubMed=7730414;
RA Steffler H.E., Ruppert C., Reinhard U., Bahler M.;
RT "A novel mammalian myosin I from rat with an SH3 domain localizes to
  cell A- and inducible, F-actin-rich structures at cell-cell contacts.";
RL J. Cell Biol. 129:819-830(1995).
CC -1- FUNCTION: Myosins are actin-based motor molecules with ATPase
  activity. Unconventional myosins serve in intracellular movements.
  Their highly divergent tails are presumed to bind to membranous
  compartments, which would be moved relative to actin filaments (By
  similarity).
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
DR EMBL; X74815; CAA52815.1; -.
DR PIR; S52517; S52517.
DR HSP; P08799; 1MDN.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD00066; SH3; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50002; SH3; 1.
KM Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
  Multigene family.
FT DOMAIN 1 679 MYOSIN HEAD-LIKE.
FT DOMAIN 695 724 IQ.
FT DOMAIN 1050 1107 SH3.
FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).
FT NP BIND 112 119 ATP (POTENTIAL).
SQ SEQUENCE 1107 AA; 12682 MW; B9D8FB8CE047148 CRC64;

```



Best Local Similarity 22.2%; Pred. No. 8.4;  
Matches 70; Conservative 43; Mismatches 114; Indels 88; Gaps 15;

QY 47 GADCAKPKOKSELEF---KENPKALERTPTSGNSKTYGKPKHLEPSHQYRELQRYA 102  
DB 476 GADQCTLLQCTCQWQISGHEHFN-----SWMOGIIHRYAGKSYDMDGFCENRR 523  
QY 103 BEVGIFFTASGMDENAVFELHNPFFVKGSGDTNNPYLEKTKAKGFPWISGQSM 162  
DB 524 D---VLF---MD---LIHMOSSELPFK-----SIFPENIQAKKKGPTTNGSKIR-- 566  
QY 163 DTMKQVOYQVKEPLNPFCLQCTSAV-----EPQPDVNLRTVISEYQKIPPDIP 211  
DB 567 ---KQANDLVSTL-----MKCTPHYRCIKENETKPKPMWESRYKQVBYLQKENTR 617  
QY 212 IGVSGHEHGTALSVAVNALGAVLEBHITLDR---TWKSGDHSASLEPGEALHVSRYR 267  
DB 618 VRRAGT-----AYRRVFOKFLQRYALITKATPFWGDKQCVL-----HILQSVN 663  
QY 268 IVERALGSPFKQPLPCENACNEKLGKSVYAKVIEGIIITMDMLTWKGEKAVPPEDI 327  
DB 664 M-----DSDQF-----QLGSKVFE-IKAPESLFLLEMEERKTKGYARVYQKTW 706  
QY 328 ENLNGKVLTYVEED 342  
DB 707 RKFVARKKYVQWRED 721  
RESULT 13  
BGLI\_BACSU STANDARD; PRT; 469 AA.  
ID BGLI\_BACSU STANDARD; PRT; 469 AA.  
AC P40740; O32287;  
AD 01-FEB-1995 (Rel. 31, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DR 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Beta-glucosidase (EC 3.2.1.21) (Cellulobiose) (Beta-D-glucoside glucohydrolase) (Amygdalase).  
GN BGLH OR N17D OR BSU39260.  
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / Marburg;  
RX MEDLINE=95189730; PubMed=7883710;  
RA Le Coq D., Lindecker C., Krueger S., Steinmetz M., Stuelke J.,  
RT "New beta-glucoside (Bgl) genes in Bacillus subtilis: the bglp gene  
RT product has both transport and regulatory functions similar to those  
RT of BglP, its Escherichia coli homologue."  
RL J. Bacteriol. 177:1527-1535(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / BGSCL1;  
RX MEDLINE=95213088; PubMed=7704263;  
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.,  
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis  
RT genome containing the hut and wpa loci."  
RL Microbiology 141:337-343(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst P., Ogasawara N., Mosser J., Albertini A.M., Alloni G.,  
RA Azevedo V., Ogasawara N., Mosser J., Albertini A.M., Alloni G.,  
RA Bourrier L., Bourrier L., Brans A., Brann M., Bignon S.C., Bron S.,  
RA Bruller S., Bruch C.V., Caldwell B., Capiano V., Carter N.M.,  
RA Choi S.K., Codani J.U., Conterton I.F., Cummings N.J., Daniel R.A.,  
RA Denicot F., Devine K.M., Dusterhoft A., Enrich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Funa S., Galazzi A., Galleron N.,  
RA Grim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grand G.,  
RA Griespelt G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,  
RA Noote D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,  
RA Priesecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleicher S., Schroeter R., Scoffone P.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
RA Toesato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,  
RA Viati A., Wambut R., Wedler E., Wedler H., Wetzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis."  
RL Nature 390:249-256(1997).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
CC -1- glucose residues with release of beta-D-glucose.  
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z34526; CA84287.1; -  
CC EMBL: D31856; BA06653.1; -  
CC EMBL: D29985; BA06657.1; -  
CC EMBL: 299124; CAB15962.1; ALT\_INIT.  
CC PIR: H69593; H69593.  
CC HSSP: P1546; 1PBG.  
CC Subtilisin; BGI0935; bglH.  
CC InterPro: IPR001360; Glyco\_hydro\_1.  
CC Pfam: PF00232; Glyco\_hydro\_1; 1.  
CC PRINTS: PR00131; GLHYDRASE1.  
CC ProDom: PD00650; Glyco\_hydro\_1; 1.  
CC PROSITE: PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.  
CC PROSITE: PS00533; GLYCOSYL\_HYDROL\_F1\_2; 1.  
CC Hydroxylase; Glycosidase; Cellulose degradation; Complete proteome.  
CC ACT SITE 175 175 PROTON DONOR (POTENTIALITY).  
CC ACT SITE 368 368 NUCLEOPHILE (BY SIMILARITY).  
CC CONFLICT 402 402 A -> R (IN REF. 1).  
CC SEQUENCE 469 AA; 53289 MW; 40C0A25F64C1C1C1 CRC64;  
Query Match 5.3%; Score 100.5; DB 1; Length 469;  
Best Local Similarity 22.2%; Pred. No. 3.2;  
Matches 72; Conservative 50; Mismatches 82; Indels 121; Gaps 20;  
QY 79 KTYG-EKRRHLEPSHOYR-LQRYAEYGFITASGMDENAVFELHNPV-----PEF 130  
DB 138 KNYGGMKRYKIEF-YERYAKTVFRKYOHKXYWT-----ENEINVLHAP- 184  
QY 131 KVGSGDTNNPYLEKTKAKGFPWISGQSMNDYKQV--YOYV-----PLN 176  
DB 185 -TGCG-----LVREGEKTLNMYQAHQIFVASALAVKAGHDIIIDS 226  
QY 177 PNFCELOCTSAVP--QPEDVNLRTVISEYQKL-PPDIP-GLYSGH-----ETGIALSV 225  
DB 227 KIGCIATNTTYPTMSKEDVYFAMENERTKILFESDVGARQAYPGYMKRYALENNIIIZEM 286  
QY 226 AAVNALGAVLEBHITLDTWKGSDH---SASLEGEALHVSRYRIVERALGSPFKQPL 281  
DB 287 AS--GDELLAHEHVD--YIGFSTYMSMASTDEELA----- 320  
QY 282 PCENACNEKLGKSVYAKV-----IEGIIITMDMLTWKGEKAVPPEDI 328

Thu May 13 11:53:10 2004

us-09-930-440b-6.rsp

Page 10

D2 321 -----KSGNLLGVYKPNYLLKSSNGWQIDPKGRLITLTYDYKPF-LPIVENGL 371  
QY 329 NLVGGKVLVVEEDPTIMEELVDNH 353  
DB 372 GAVDK-----VBEODTODDPRINY 391

RESULT 14

MYO\_HUMAN STANDARD; PRT; 1109 AA.  
ID MYO\_HUMAN  
AC Q12965;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Myosin Ic (Myosin Ic).  
GN MYOIC OR MYOIC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OK NCBI\_Taxid=9606;  
RN (1)  
RP SEQUENCE OF 1-245 FROM N.A.  
RX MEDLINE=95018277; PubMed=7932763;  
RA Bement W.M., Wirth J.A., Mooseker M.S.;  
RT "Cloning and mRNA expression of human unconventional myosin-IC. A  
RT homologue of amoeboid myosins-I with a single IQ motif and an SH3  
RT domain";  
RL J. Mol. Biol. 243:356-363(1994).  
CC -1- FUNCTION: Myosins are actin-based motor molecules with ATPase  
CC activity. Unconventional myosins serve in intracellular movements.  
CC Their highly divergent tails are presumed to bind to membranous  
CC compartments, which would be moved relative to actin filaments (By  
CC similarity).  
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.  
CC -1- SIMILARITY: Contains 1 IQ domain.  
CC -1- SIMILARITY: Contains 1 SH3 domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
CC -----  
DR EMBL: U14391; AAA62667.1; -  
DR PIR: S53601; S53601.  
DR KSSP: P08799; IAMD.  
DR Genew: HGNC:1599; MYOIE.  
DR MIM: 601479; -  
DR GO: GO:0015629; C:actin cytoskeleton; TAS.  
DR GO: GO:0003774; F:motor activity; TAS.  
DR GO: GO:0008570; F:myosin ATPase activity; TAS.  
DR InterPro: IPR000048; IQ region.  
DR InterPro: IPR001609; myosin\_head.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00612; IQ; 1.  
DR Pfam: PF00663; myosin\_head; 1.  
DR Pfam: PF00018; SH3; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR ProDom: PD000355; myosin\_head; 1.  
DR ProDom: PD000066; SH3; 1.  
DR SMART: SM00242; MYSC; 1.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS50096; IQ; 1.  
DR PROSITE: PS50002; SH3; 1.  
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;  
KW Multigene family.  
FT DOMAIN 1 679 MYOSIN HEAD-LIKE.  
FT DOMAIN 695 724 IQ.  
FT DOMAIN 1052 1109 SH3.  
FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).

FT NP BIND 112 119 ATP (POTENTIAL).  
SQ SEQUENCE 1109 AA; 127040 MW; 0C1C74BA3705905 CRC64;  
Query Match 5.3%; Score 100; DB 1; Length 1109;  
Best Local Similarity 22.5%; Pred. No. 10;  
Matches 64; Conservative 41; Mismatches 91; Indels 88; Gaps 15;

QY 47 GADPAKQKSELEF-----KFKRALEBPYTSKSKMGKTYGHRHLEFSHDQYRELQRYA 102  
DB 476 GADDTLLQKQOMQSGHEHN-----SWNGQFIHHYAGVSYDMQFCRRNR 523  
QY 103 EEWGIFPTASGDMENAVELHNLVPPFKYSGDTPNPPLEKTKAKGRPMVSSGQSM 162  
DB 524 D---VLF---MD--LIELMQSELPIK-----SLFPEMLQADKKGRPTTAGSKTK-- 566  
QY 163 DTMKQVQVYKPLPNFCPLQCTSAV-----PLQPEVNLRLVISEYOKLPPDIP 211  
DB 567 ---KQANDLVSTL-----MKCTPHYTRCKPKNETKPRPMEERVRHGVYELGKENIR 617  
QY 212 IGVSGETGIAISVAVALGAKVLERHITDK---TWKSDHSASLEPGLAEVRSVR 267  
DB 618 VRAGY-----AYRIRFKFQRYAILTKATPMPQGEKQGVL-----HLLQSVN 663  
QY 268 LVERALSPPTQQLPCEMACNEKKGKSVAKVKIPEGLITLMDM 311  
DB 664 M-----DSDQF-----QGRSKVP-ITAPESFLLEEM 690

RESULT 15

GPR\_BACME STANDARD; PRT; 371 AA.  
ID GPR\_BACME  
AC P22321;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Germination protease precursor (EC 3.4.24.78) (Spore protease) (GPR  
DE endopeptidase) (Germination proteinase).  
GN GPR.  
OS Bacillus megaterium.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OK NCBI\_Taxid=1404;  
RN (1)  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-31.  
RX STRAIN=ATCC 12872 / OMDC B1551;  
RX MEDLINE=91100296; PubMed=1840582;  
RA Suseman M.D., Setlow P.;  
RT "Cloning, nucleotide sequence, and regulation of the Bacillus  
RT subtilis gpr gene, which codes for the protease that initiates  
RT degradation of small, acid-soluble proteins during spore  
RT germination";  
RL J. Bacteriol. 173:291-300(1991).  
RN (2)  
RP CHARACTERIZATION.  
RX MEDLINE=98422459; PubMed=9748439;  
RA Nessi C., Jedrzejas M.J., Setlow P.;  
RT "Structure and mechanism of action of the protease that degrades  
RT small, acid-soluble spore proteins during germination of spores of  
RT Bacillus species";  
RL J. Bacteriol. 180:5077-5084(1998).  
RN (3)  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF ZYMAGEN P46.  
RX MEDLINE=20323395; PubMed=1864493;  
RA Ponuraj K., Rowland S., Nessi C., Setlow P., Jedrzejas M.J.;  
RT "Crystal structure of a novel germination protease from spores of  
RT Bacillus megaterium: structural arrangement and zymogen activation";  
RL J. Mol. Biol. 300:11-10(2000).  
CC -1- FUNCTION: Initiates the rapid degradation of small, acid-soluble  
CC proteins during spore germination.  
CC -1- CATALYTIC ACTIVITY: Endopeptidase action with P4 Glu or Asp, P1  
CC preferably Glu > Asp, P1' hydrophobic and P2' Ala.  
CC -1- SUBUNIT: Homotetramer.  
CC -1- DEVELOPMENTAL STAGE: GPR transcription occurs during sporulation  
CC in forespore first by sigma-F and then by sigma-G.

```
CC -1- P1M: Autoproteolytically processed. The inactive tetrameric
CC zymogen termed p46 autoprocesses to a smaller form termed p41,
CC which is active only during spore germination.
CC -1- SIMILARITY: Belongs to peptidase family M63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M55262; AAA22499.1; -
CC PIR, A39198; A39198.
CC PDB, 1C8B; 03-MAY-01.
CC MEROPS; M63.001; -.
CC HAMAP; MF_00626; -; 1.
CC InterPro; IPR005080; Peptidase_M63.
CC Pfam; PF03418; Peptidase_U3; 1.
CC ProDom; PD041835; Peptidase_M63; 1.
CC TIGRFAMs; TIGR01441; GPR; 1.
CC HydroLase; Protease; Zymogen; 3D-structure.
CC PROPEP 1 15
CC FT CHAIN 16 371 GERMINATION PROTEASE.
CC SO SEQUENCE 371 AA; 40626 MW; 40AB862ECA2F835D CRC64;

Query Match 5.28; Score 98.5; DB 1; Length 371;
Best Local Similarity 20.98; Pred. No. 3.4;
Matches 63; Conservative 44; Mismatches 120; Indels 75; Gaps 12;

QY 100 RYA--EEVGIFFPASGMDENAVE-----FLHELAVP---FFKVGSGDTNFP--- 141
DB 71 RYVTLSEVGIRBQDTEKQEBMEBEVFPKELNPFKSLNIPDDASCLVVGIMLSTPDL 130
QY 142 -----YLEKTAKGRPM--VISS--GMQSDPTMKOVQIVKLNPNFC 180
DB 131 GPKAVDNLITRHLFELQPSVCDGFRFVSAIVGVGAKTGIIETSDIIFGVKKVNDPFI 190
QY 181 F-IQCTSAVLPQPDVNLAVISEYQKLPDIPIGYSGHETG---IAISVAVALGAKYLE 236
DB 191 IAIIDALARST--ERVNATIQISDSGIHPSGVGNKRKEISYETLGIPTVAIGIPYVDA 248
QY 237 RHITLDKTMKSDHSASLEPGEIAIVRSVRLVERALGSPTKOLLPCEMAENELGKSVV 296
DB 249 VSIITSD-----TIDFILKHFGRKMKQCKPSKSLPFGMTGEG----- 286
QY 297 AKVYIPSGEILTMQMLTVKVGEPKAYPPEDINLV-----GKKVAVTEEDDTIMEEL 349
DB 287 -KKKLTEDDLANEBQRTYLGMLGTLPEDEKRRLIHEVLAPLGNELVVTPKEVDMFTEDM 345
QY 350 VD 351
DB 346 AN 347
```

Search completed: May 6, 2004, 09:05:03  
Job time : 12.4405 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:19 ; Search time 35.1395 Seconds

(without alignments)  
3223.466 Million cell updates/sec

Title: US-09-930-440B-6  
Perfect score: 1887  
Sequence: 1 MPELELCPGRWVGQHPF.....EEDTIMELVNHNKIKS 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1810	95.9	359	11	Q99J77 mus musculi
2	1804	95.6	359	11	Q99J77 mus musculi
3	854	45.3	372	5	Q9V674 drosophila
4	547.5	29.0	280	16	Q8K655 chlorobium
5	534.5	28.3	341	2	Q87186 streptococc
6	534.5	28.3	341	16	Q8E501 streptococc
7	532.5	28.2	339	2	Q9ALM6 galax streptococc
8	532.5	28.2	339	2	Q93TT2 galax streptococc
9	532.5	28.2	341	2	Q9RPO0 galax streptococc
10	532.5	28.2	341	16	Q8D2B3 galax streptococc
11	526	27.9	749	16	Q82UC3 galax streptococc
12	524	27.8	311	16	Q82HY4 galax streptococc
13	518	27.5	334	2	Q7X523 campylobact
14	516.5	27.4	333	16	Q8DD27 vibrio vuln
15	514	27.2	361	2	Q8KX22 pseudomonas
16	512	27.1	312	16	Q9AK45 streptomyces

17	510.5	27.1	346	2	Q46675
18	510	27.0	338	2	Q9RDX5
19	509.5	27.0	351	16	Q89J06
20	507.5	26.9	753	16	Q8E504
21	506.5	26.8	341	2	Q9A016
22	500.5	26.5	641	16	Q89HL9
23	497	26.3	754	16	Q7U964
24	487	25.8	334	16	Q9PMX2
25	472	25.0	333	16	Q7U911
26	472	25.0	346	2	Q93N02
27	461.5	24.5	356	16	Q89HL8
28	437.5	23.2	286	17	Q8TUL5
29	435.5	23.1	344	16	Q8F508
30	435	23.1	346	16	Q8A711
31	431.5	22.9	357	16	Q87T70
32	419	22.2	350	16	Q97H26
33	410	21.7	351	2	Q8KH52
34	393.5	20.9	352	2	Q9R9S2
35	390.5	20.7	351	16	Q8R1B8
36	384.5	20.4	351	2	Q8KH54
37	382.5	20.3	349	16	Q57265
38	381.5	20.2	332	16	Q8F330
39	374.5	19.8	354	16	Q7V953
40	367.5	19.5	343	2	Q93J78
41	363.5	19.3	354	2	Q939M1
42	360.5	19.1	229	2	Q9R9F4
43	358	19.0	356	16	Q9A4G6
44	356.5	18.9	343	16	Q9PMY2
45	356	18.9	343	2	Q9EU02

## ALIGNMENTS

RESULT 1  
Q99J77 PRELIMINARY; PRT; 359 AA.  
ID Q99J77;  
AC 01-JUN-2001 (TREMBL) 17, Created)  
DT 01-JUN-2001 (TREMBL) 17, Last sequence update)  
DE 01-OCT-2003 (TREMBL) 25, Last annotation update)  
DE Similar to N-acetylneuraminic acid phosphate synthase, sialic acid synthase.  
GN NANS OR 463241804RIK OR SAS.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RA "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
EMBL: BC003507; AA003507.1; -;  
EMBL: AK076290; BA036290.1; -;  
HSP: P19614; JUIA.  
DR MGD; MGI:2149820; Nans.  
DR GO; GO:0005829; C:cytosol; IDA.  
DR GO; GO:0019007; P:N-acetylneuraminic acid phosphate synthase . . . ; IDA.  
DR InterPro; IPR006013; AntifreezeZell.  
DR InterPro; IPR006014; AntifreezeZell.  
DR InterPro; IPR006190; AntifreezeZell.  
DR InterPro; IPR004144; Neut.  
DR Pfam; PF01354; AntifreezeZell; 1.  
DR Pfam; PF03102; Neut; 1.

DR PRINTS; PRO0357; ANTIFREEZII.  
 DR PRODOM; PD003258; AntifreezeII; 1.  
 DR PROSITE; PS50844; AFP LIKE; 1.  
 SQ SEQUENCE 359 AA; 40024 MW; 4C66CB883558A373 CRC64;

Query Match 95.9%; Score 1810; DB 11; Length 359;  
 Best Local Similarity 94.4%; Pred. No. 8.2e-141;  
 Matches 339; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MPELELCGRWVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 60  
 DB 1 MPELELCGRWVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 60  
 QY 61 KFNKRALERPYSKSHSGKTYGKHLEPESHQYRELQKVAEYVGIPTASGMDENAVE 120  
 DB 61 KFNKRALERPYSKSHSGKTYGKHLEPESHQYRELQKVAEYVGIPTASGMDENAVE 120  
 QY 121 FLHELVNPFKVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 180  
 DB 121 FLHELVNPFKVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 180  
 QY 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDIPITGSGHETGIALSYAAVALGAKVLEHIT 240  
 DB 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDIPITGSGHETGIALSYAAVALGAKVLEHIT 240  
 QY 241 LDKTWGSDBSASLEPGELEALVRSVRLVERALGSPTRKOLPCEMACNEKLGKSVAVKX 300  
 DB 241 LDKTWGSDBSASLEPGELEALVRSVRLVERALGSPTRKOLPCEMACNEKLGKSVAVKX 300  
 QY 301 IPEGTILFMMLTVKVGEPKAYPPEDINLVGKXVLTVEEDDTIMEELVDNKGKTKS 359  
 DB 301 IPEGTILFMMLTVKVGEPKAYPPEDINLVGKXVLTVEEDDTIMEELVDNKGKTKS 359

## RESULT 2

Q9JUH0 PRELIMINARY; PRT; 359 AA.  
 AC Q9JUH0; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE N-acetylneuraminic acid 9-phosphate synthetase.  
 GN NANS OR SAS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Sutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20334323; PubMed=10873558;  
 RA Nakata D., Close B.E., Colley K.J., Matsuda T., Kitajima K.;  
 RT "Molecular cloning and expression of the mouse N-acetylneuraminic acid  
 RT 9-phosphate synthase which has not the deammonneuraminic acid (KDN)  
 RT 9-phosphate synthase activity";  
 RL Biochem Biophys Res Commun. 273:642-648(2000).  
 DR EMBL; AB041263; BA98131.1; -;  
 DR FIR; JC7321; JC7321.  
 DR HSSP; P19614; J21A.  
 DR MGD; MG1:2149820; Mans.  
 DR GO; GO:0005829; C:cytosol; IDA.  
 DR InterPro; IPR006190; Antifreeze\_1like.  
 DR InterPro; IPR004144; Neut.  
 DR Pfam; PF03102; Neut; 1.  
 DR PROSITE; PS50844; AFP\_LIKE; 1.  
 SQ SEQUENCE 359 AA; 39994 MW; 5F8F7D40C558A373 CRC64;

Query Match 95.6%; Score 1804; DB 11; Length 359;  
 Best Local Similarity 94.2%; Pred. No. 2.5e-140;  
 Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 MPELELCGRWVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 60

DB 1 MPELELCGRWVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 60  
 QY 61 KFNKRALERPYSKSHSGKTYGKHLEPESHQYRELQKVAEYVGIPTASGMDENAVE 120  
 DB 61 KFNKRALERPYSKSHSGKTYGKHLEPESHQYRELQKVAEYVGIPTASGMDENAVE 120  
 QY 121 FLHELVNPFKVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 180  
 DB 121 FLHELVNPFKVGSGDPTNNPYLEKTAAGKGRPMVSSGMSQMDTKQVYQIVKPLNPNFC 180  
 QY 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDIPITGSGHETGIALSYAAVALGAKVLEHIT 240  
 DB 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDIPITGSGHETGIALSYAAVALGAKVLEHIT 240  
 QY 241 LDKTWGSDBSASLEPGELEALVRSVRLVERALGSPTRKOLPCEMACNEKLGKSVAVKX 300  
 DB 241 LDKTWGSDBSASLEPGELEALVRSVRLVERALGSPTRKOLPCEMACNEKLGKSVAVKX 300  
 QY 301 IPEGTILFMMLTVKVGEPKAYPPEDINLVGKXVLTVEEDDTIMEELVDNKGKTKS 359  
 DB 301 IPEGTILFMMLTVKVGEPKAYPPEDINLVGKXVLTVEEDDTIMEELVDNKGKTKS 359

## RESULT 3

Q9VG74 PRELIMINARY; PRT; 372 AA.  
 AC Q9VG74; Q95VY1;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE CG5232 protein (Neu5Ac synthase).  
 GN NSU53C OR CG5232.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pelecyota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;

RP SEQUENCE FROM N.A.  
 RX STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Heit G., Andrews-Pfannkoch C., Baldwin D.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Beasley E.M.,  
 RA Bailew R.M., Baer A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bens P.V., Berman B.P., Bhattacharya D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brodeur P., Brothier P.,  
 RA Burlis K.C., Butcher M.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.B., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertzer W.S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Heston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris M.P., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclob J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svendsen R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.N., Weissbach J.,

Query Match 45.3%; Score 854; DB 5; Length 372;  
Best Local Similarity 46.8%; Pred. No. 5, 6e-62;  
Matches 167; Conservative 64; Mismatch 102; Indels 24; Gaps 3

20 FTAAIGQHOGSLDVAKKIMIMAKSGAGDKAFQKSELEFKNRKALRPTYSKSGWK 79  
16 YIIAIGQHOGSLDVAKKIMIMAKSGAGDKAFQKSELEFKNRKALRPTYSKSGWK 75

80 TTGHRKHLFSDHYRELQRYAEVGIPTFASGDMBAVEPLHNLNPFKSGSDTN 139

Db	76	TYGHHKEVLEFSDQYLQLAHGKELNVDPFASANDRSLEFLSALNVPEIKIGSSDANN	135
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV	295
Db	256	LKALTVAITPFKLSSVPMPPQEIYKCKLNGEBHELMALQHVESKTLTLPCELPCKNKAGSI	315
Qy	296	VAKXIIEGTLITMDMLTVKNGEKPAPVEDPILVNGKVLVTEBEDTIEMELVND	352
Db	316	VAAINNKGRLOLADMAIKVSEBSGLTAEDFDVLQKLELADNGIDEPBILGNSIIN	372

  

Db	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV	295
Db	256	LKALTVAITPFKLSSVPMPPQEIYKCKLNGEBHELMALQHVESKTLTLPCELPCKNKAGSI	315
Qy	296	VAKXIIEGTLITMDMLTVKNGEKPAPVEDPILVNGKVLVTEBEDTIEMELVND	352
Db	316	VAAINNKGRLOLADMAIKVSEBSGLTAEDFDVLQKLELADNGIDEPBILGNSIIN	372

  

Db	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV	295
Db	256	LKALTVAITPFKLSSVPMPPQEIYKCKLNGEBHELMALQHVESKTLTLPCELPCKNKAGSI	315
Qy	296	VAKXIIEGTLITMDMLTVKNGEKPAPVEDPILVNGKVLVTEBEDTIEMELVND	352
Db	316	VAAINNKGRLOLADMAIKVSEBSGLTAEDFDVLQKLELADNGIDEPBILGNSIIN	372

  

Db	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV	295
Db	256	LKALTVAITPFKLSSVPMPPQEIYKCKLNGEBHELMALQHVESKTLTLPCELPCKNKAGSI	315
Qy	296	VAKXIIEGTLITMDMLTVKNGEKPAPVEDPILVNGKVLVTEBEDTIEMELVND	352
Db	316	VAAINNKGRLOLADMAIKVSEBSGLTAEDFDVLQKLELADNGIDEPBILGNSIIN	372

  

Db	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV	295
Db	256	LKALTVAITPFKLSSVPMPPQEIYKCKLNGEBHELMALQHVESKTLTLPCELPCKNKAGSI	315
Qy	296	VAKXIIEGTLITMDMLTVKNGEKPAPVEDPILVNGKVLVTEBEDTIEMELVND	352
Db	316	VAAINNKGRLOLADMAIKVSEBSGLTAEDFDVLQKLELADNGIDEPBILGNSIIN	372

  

Db	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV	295
Db	256	LKALTVAITPFKLSSVPMPPQEIYKCKLNGEBHELMALQHVESKTLTLPCELPCKNKAGSI	315
Qy	296	VAKXIIEGTLITMDMLTVKNGEKPAPVEDPILVNGKVLVTEBEDTIEMELVND	352
Db	316	VAAINNKGRLOLADMAIKVSEBSGLTAEDFDVLQKLELADNGIDEPBILGNSIIN	372

  

Db	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Qy	140	PPYLKTKAKGKPMWLTSSGMSMDTKQVQYVKEPL-NENPFCLOCTSAVELQPEDVNLK	198
Db	136	PPLLKTKAANLNLPLVLTSTGQTMQVTRIVQVMESEKREYALMGVSSIFPTDKDSLQ	195
Qy	199	VISEYQCLPDPIDPIGSGSHETGIAISVAVALGAKVLEHRIITLDKTKWGSDBSASLEPGE	258
Db	196	LISVLRFRFRVNAIGSGHLEGVILISQAVALGARIVYRHPTLDKSGQSDHRCSTLEPQ	255
Qy	259	LAELVRSV-----RVERALGS-----PTKQLVCEWACNEKLGKV</	

QY 249 DHASLEPGLAEIVRSVRLVRAIGSPETKOLLPCEMACNEKI 291  
 DB 235 DQASVEISGMSRLVSNIRDLERKALGDGVKRYDGEAAARKKL 277

## RESULT 5

PRELIMINARY: PRT: 341 AA.

AC 087186 PRELIMINARY: PRT: 341 AA.  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE NEUB protein.  
 GN NEUB.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=1311;  
 [1]  
 [1]  
 RA Yamaoka S., Miyake K., Iijima S.;  
 RT "Identification and Characterization of cps (capsular polysaccharide)  
 RT Genes from Streptococcus agalactiae Type Ia.";  
 RT Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Miyake K., Yamaoka S., Koike Y., Matanabe M., Iijima S.;  
 RT "Molecular Characterization of Type-Specific Capsular Polysaccharide  
 RT Biosynthesis Genes of Streptococcus agalactiae Type Ia.";  
 RT Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB017355; BAA8287.1; -;  
 DR EMBL: AB028896; BAA8287.1; -;  
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro: IPR006014; Antifreeze dom.  
 DR InterPro: IPR006190; Antifreeze like.  
 DR InterPro: IPR004144; Neub.  
 DR Pfam: PF01354; Antifreeze; 1.  
 DR Pfam: PF03102; Neub; 1.  
 DR PROSITE: PS50844; AFP\_LIKE; 1.  
 DR PROSITE: PS50844; AFP\_LIKE; 1.  
 SQ SEQUENCE 341 AA; 38091 MW; 17620BF82A2EA340 CRC64;

Query Match 28.3%; Score 534.5; DB 2; Length 341;  
 Best Local Similarity 36.8%; Pred. No. 1.1e-35;  
 Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FTIAEIGNHGDLDVAKMIRMAKCGADCAKFOKSELEFKRKALERTPTYSKSHGK 79  
 DB 3 YIIAEIGNHGDLDVAKMIRMAKCGADCAKFOKSELEFKRKALERTPTYSKSHGK 61  
 QY 80 TYG--EKKHLEPESHDOYRELORVAREVGIFPTASGDMENAVBELHNLVPPFVSGSDT 137  
 DB 62 ADSOLEMTKRLSELEPESHDOYRELORVAREVGIFPTASGDMENAVBELHNLVPPFVSGSDT 121  
 QY 138 NNPPYLEKTAKKGRPMYISSGQSDMTKQYQIVKPLNP----FCLQCTSAVPLQPE 193  
 DB 122 TNLPLEKIGKQCKVILSTGMA---VMEIHQAVNILRQGTDISILHCTTEYPTPY 178  
 QY 194 DVNLRISEYQKLPDIPIGSGHETGIALISVAVALGAKYLERHITLDTKWKSDSHAS 253  
 DB 179 SLNLNVHITLDEKFDLTIGSDHSIGSEVPIIAAANGAEVIEKHFTLDTNMGBDHSAS 238  
 QY 254 LEPGLAEIVRSVRLVRAIGSPETKOLLPCEMACNEKIGSVAKYKIPESITLMDMLT 313  
 DB 239 ATPDILALVKGRIEVALGRPEKIPPEVEKKKIYARKSVVALKPIKGDIVSIENIT 298  
 QY 314 VKVGEF-KAYPPEDIENLVGKKVLTVEEDTITBELVDN 352  
 DB 299 VK--RFGNGISPMNMYDILGOEADDFEDEVIRDSRFEN 336

RESULT 6  
 Q8E501

ID Q8E501 PRELIMINARY: PRT: 341 AA.  
 AC Q8E501;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN NEUB OR GBS1236.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=216495;  
 [1]  
 [1]  
 RA SEQUENCE FROM N.A.  
 RC STEIN-NEUB316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusnok C., Buchrieser C., Chevallier F., Frangoul L.,  
 RA Masdek T., Zouine M., Couve E., Lailoui L., Poyart C., Tilleu-Quot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RT Mol. Microbiol. 45:1498-1513(2002).  
 DR EMBL: AL766849; CAD46895.1; -;  
 DR SsalIst; gbs1236; -;  
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro: IPR006014; Antifreeze dom.  
 DR InterPro: IPR006190; Antifreeze like.  
 DR InterPro: IPR004144; Neub.  
 DR Pfam: PF01354; Antifreeze; 1.  
 DR Pfam: PF03102; Neub; 1.  
 DR PROSITE: PS50844; AFP\_LIKE; 1.  
 DR PROSITE: PS50844; AFP\_LIKE; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 341 AA; 38091 MW; 17620BF82A2EA340 CRC64;

Query Match 28.3%; Score 534.5; DB 16; Length 341;  
 Best Local Similarity 36.8%; Pred. No. 1.1e-35;  
 Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FTIAEIGNHGDLDVAKMIRMAKCGADCAKFOKSELEFKRKALERTPTYSKSHGK 79  
 DB 3 YIIAEIGNHGDLDVAKMIRMAKCGADCAKFOKSELEFKRKALERTPTYSKSHGK 61  
 QY 80 TYG--EKKHLEPESHDOYRELORVAREVGIFPTASGDMENAVBELHNLVPPFVSGSDT 137  
 DB 62 ADSOLEMTKRLSELEPESHDOYRELORVAREVGIFPTASGDMENAVBELHNLVPPFVSGSDT 121  
 QY 138 NNPPYLEKTAKKGRPMYISSGQSDMTKQYQIVKPLNP----FCLQCTSAVPLQPE 193  
 DB 122 TNLPLEKIGKQCKVILSTGMA---VMEIHQAVNILRQGTDISILHCTTEYPTPY 178  
 QY 194 DVNLRISEYQKLPDIPIGSGHETGIALISVAVALGAKYLERHITLDTKWKSDSHAS 253  
 DB 179 SLNLNVHITLDEKFDLTIGSDHSIGSEVPIIAAANGAEVIEKHFTLDTNMGBDHSAS 238  
 QY 254 LEPGLAEIVRSVRLVRAIGSPETKOLLPCEMACNEKIGSVAKYKIPESITLMDMLT 313  
 DB 239 ATPDILALVKGRIEVALGRPEKIPPEVEKKKIYARKSVVALKPIKGDIVSIENIT 298  
 QY 314 VKVGEF-KAYPPEDIENLVGKKVLTVEEDTITBELVDN 352  
 DB 299 VK--RFGNGISPMNMYDILGOEADDFEDEVIRDSRFEN 336

## RESULT 7

Q9ALM6 PRELIMINARY: PRT: 339 AA.

AC Q9ALM6;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE Putative N-acetyl neuraminic acid synthetase Neub.  
 GN NEUB.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

```

OC Streptococcus.
OX NCBI_TaxID=1311;
RN (1)
R2 SEQUENCE FROM N.A.
RC STRAIN=NT6;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF337958; AKL1670.1;
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze_dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; Neut.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; Neut; 1.
DR PROSITE; PS50844; AFP_LIKE; 1.
SQ SEQUENCE 339 AA; 3773 MW; 156C2DBBF1D453C7 CRC64;

Query Match 28.2%; Score 532.5; DB 2; Length 339;
Best Local Similarity 36.8%; Pred. No. 1.6e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FIIAIGNHQGDLDVAKMIRAKKCGADCAKFOKSELEFKENKALERTPTSGHSGK 79
DB 3 YIIAIGNHGNDIILAKMVDVAVSCGVDAVKQTFKAEKISKAPAEY-QATATGT 61
QY 80 TYG--EHKRLHPSHDQYRELQRYAEVGIFFTASGMDMAVEFLHETLVPEFKVSGDT 137
DB 62 ADSQLEMTKRLBLESEELTMDYALSKGVETSTPFDESELEFLISTMDPIYKIPSEI 121
QY 138 NNFPYLEKTKAKGRPMVISGQMSMDTMQVQIVKPLNPN---FCFLQCTSAVPLQPE 193
DB 122 TNLPLEKIGKQOKKVIILSTGMA--VMEIHOAVNIIIRONGTDTDISILHCTTEYPTFP 178
QY 194 DYNLAVISEYQKLPDIPIGYSGHETGIAISVAVALGAKVLRHITLDTKMGSDHSAS 253
DB 179 SLINLVNHTLKDPEFDLTIGYSDHSIGSEVPIAAALGAEVIEKHTLDTMNEGPHKAS 238
QY 254 LEPGEIAELVSVRLVERALASPTKQLPCENACNEKLGKSVAVAKVPIEGTILTMMDLT 313
DB 239 ATPDLIALLVKGVRIVEQALGRFEKIPDPVEKKNIVARKSVVALKPIKKGDIYSIENT 298
QY 314 VKVGEF-KAYPPEDIENLVGKKVLTVEEDDTIMEBLVDN 352
DB 299 VK--RPGNGISPMNMYDILQGEADDFEBDEVIDRSEFN 336

RESULT 8
Q937T2 PRELIMINARY; PRT; 339 AA.
ID 0937T2
AC 0937T2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative N-acetyl neuraminic acid synthetase Neut.
GN NEUT.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN (1)
R2 SEQUENCE FROM N.A.
RC STRAIN=CMTC 1/82;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355776; AAK43615.1;
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze_dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; Neut.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; Neut; 1.
DR PROSITE; PS50844; AFP_LIKE; 1.
SQ SEQUENCE 339 AA; 37791 MW; 156C2B2BF31457C7 CRC64;

```

```

Query Match 28.2%; Score 532.5; DB 2; Length 339;
Best Local Similarity 36.8%; Pred. No. 1.6e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FIIAIGNHQGDLDVAKMIRAKKCGADCAKFOKSELEFKENKALERTPTSGHSGK 79
DB 3 YIIAIGNHGNDIILAKMVDVAVSCGVDAVKQTFKAEKISKAPAEY-QATATGT 61
QY 80 TYG--EHKRLHPSHDQYRELQRYAEVGIFFTASGMDMAVEFLHETLVPEFKVSGDT 137
DB 62 ADSQLEMTKRLBLESEELTMDYALSKGVETSTPFDESELEFLISTMDPIYKIPSEI 121
QY 138 NNFPYLEKTKAKGRPMVISGQMSMDTMQVQIVKPLNPN---FCFLQCTSAVPLQPE 193
DB 122 TNLPLEKIGKQOKKVIILSTGMA--VMEIHOAVNIIIRONGTDTDISILHCTTEYPTFP 178
QY 194 DYNLAVISEYQKLPDIPIGYSGHETGIAISVAVALGAKVLRHITLDTKMGSDHSAS 253
DB 179 SLINLVNHTLKDPEFDLTIGYSDHSIGSEVPIAAALGAEVIEKHTLDTMNEGPHKAS 238
QY 254 LEPGEIAELVSVRLVERALASPTKQLPCENACNEKLGKSVAVAKVPIEGTILTMMDLT 313
DB 239 ATPDLIALLVKGVRIVEQALGRFEKIPDPVEKKNIVARKSVVALKPIKKGDIYSIENT 298
QY 314 VKVGEF-KAYPPEDIENLVGKKVLTVEEDDTIMEBLVDN 352
DB 299 VK--RPGNGISPMNMYDILQGEADDFEBDEVIDRSEFN 336

RESULT 9
Q937C0 PRELIMINARY; PRT; 341 AA.
ID 0937C0
AC 0937C0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CPSN (NEUT).
GN CPSN (NEUT).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN (1)
R2 SEQUENCE FROM N.A.
RC STRAIN=COH1;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF163833; AAD53074.1;
DR EMBL; AF349539; AAK29661.1;
DR PIR; T44651; T44651.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze_dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; Neut.

```



Thu May 13 11:53:10 2004

us-09-930-440b-6.rpt

Page 6

DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF03102; Neut; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
SQ SEQUENCE 341 AA; 38033 MW; 148B756C2B2BF314 CRC64;  
Query Match: 28.2%; Score 532.5; DB 2; Length 341;  
Best Local Similarity 36.8%; Pred. No. 1.6e-35;  
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;  
QY 20 PIIAETGQNHQGDLDVAKRMIMARECGADCAKFOKSELEFKENKALERPYSKSGWK 79  
DB 3 YIIAETGQNHQGDLDVAKRMIMARECGADCAKFOKSELEFKENKALERPYSKSGWK 61  
QY 80 TYG--EHRKLEPSHDQRELQRYAEVGIFFITASGNDMAVEFIHELVNPFVSGGDT 137  
DB 62 ADSQLEMTKRLSLFEEYLEMRDVALSKGVETSTPDEESLEFLISTDMPYKIPSGEI 121  
QY 138 NNPFYLEKTKAKGRPMVTSGGMSMDTMKQVQIVKPLNPN---FCFLQCTSAVPLQPE 193  
DB 122 TNLPILEKIGKQKQKVLSTGMA---VMEIHQAVNILLRQNGTTDISILHCTTEYPTYP 178  
QY 194 DVNLKRVISEYQKLPDIPIGYSGHETGIAISVAVALGAKYLERHTLIDKTKWSDHSAS 253  
DB 179 SLINLVNHTLTKDEFKDLITIGYSDHSIGSEVPDIAAAMGAELVIEKHFTLDTNMEGPDHKS 238  
QY 254 LEPGELAEIVRSVLRALGSPTRKOLLPCENACNEKLGKSVAAKVIPEGTILTMMLT 313  
DB 239 ATPDILALVKGRIEVLQGRFEKIPDVEEKNNIVAKSVVALLKPKIGDIYSIENIT 298  
QY 314 VKVGEF-KAYPPEIDIFNLVGRKVLVTVSEDDTMEELVDN 352  
DB 299 VK--RPGNGISPMNWYDILGQEAQDDFEDEVIRDSRFEN 336  
RESULT 10  
Q8DZ3 PRELIMINARY; PRT; 341 AA.  
AC Q8DZ3;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-JUN-2003 (Tremblrel. 23, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE N-acetyl neuraminic acid synthetase Neut.  
GN NEUB OR SAG1161.  
OS Streptococcus agalactiae (serotype V).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OC NCBI\_TaxID=216466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2603 V/R / Serotype V.  
RX MEDLINE=22222988; PubMed=12200547;  
RA Tettelin H., Nisigian V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
RA Messels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,  
RA DeBoy R.T., Dackiw A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
RA Paduane D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
RA Carry H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
RA Iacobini E.T., Brelton C., Galli G., Mariani M., Vegni F., Malone D.,  
RA Rinaldo D., Rappunli R., Telford J.L., Kaeper D.L., Grandi G.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative genomic analysis of an  
RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
DR EMBL: AB014245; J04651.  
DR PIR: T4651; T44651.  
DR TIGR: SAG1161;  
DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro: IPR006014; Antifreeze\_dom.  
DR InterPro: IPR006190; Antifreeze\_like.  
DR InterPro: IPR004144; Neut.  
DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF03102; Neut; 1.  
DR Pfam: PF03102; Neut; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.

KM Complete proteome.  
SQ SEQUENCE 341 AA; 38033 MW; 148B756C2B2BF314 CRC64;  
Query Match: 28.2%; Score 532.5; DB 16; Length 341;  
Best Local Similarity 36.8%; Pred. No. 1.6e-35;  
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;  
QY 20 PIIAETGQNHQGDLDVAKRMIMARECGADCAKFOKSELEFKENKALERPYSKSGWK 79  
DB 3 YIIAETGQNHQGDLDVAKRMIMARECGADCAKFOKSELEFKENKALERPYSKSGWK 61  
QY 80 TYG--EHRKLEPSHDQRELQRYAEVGIFFITASGNDMAVEFIHELVNPFVSGGDT 137  
DB 62 ADSQLEMTKRLSLFEEYLEMRDVALSKGVETSTPDEESLEFLISTDMPYKIPSGEI 121  
QY 138 NNPFYLEKTKAKGRPMVTSGGMSMDTMKQVQIVKPLNPN---FCFLQCTSAVPLQPE 193  
DB 122 TNLPILEKIGKQKQKVLSTGMA---VMEIHQAVNILLRQNGTTDISILHCTTEYPTYP 178  
QY 194 DVNLKRVISEYQKLPDIPIGYSGHETGIAISVAVALGAKYLERHTLIDKTKWSDHSAS 253  
DB 179 SLINLVNHTLTKDEFKDLITIGYSDHSIGSEVPDIAAAMGAELVIEKHFTLDTNMEGPDHKS 238  
QY 254 LEPGELAEIVRSVLRALGSPTRKOLLPCENACNEKLGKSVAAKVIPEGTILTMMLT 313  
DB 239 ATPDILALVKGRIEVLQGRFEKIPDVEEKNNIVAKSVVALLKPKIGDIYSIENIT 298  
QY 314 VKVGEF-KAYPPEIDIFNLVGRKVLVTVSEDDTMEELVDN 352  
DB 299 VK--RPGNGISPMNWYDILGQEAQDDFEDEVIRDSRFEN 336  
RESULT 11  
Q82UC3 PRELIMINARY; PRT; 749 AA.  
AC Q82UC3;  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Type III antifreeze protein: CBS domain: Neut family (BC 4.1.3.-).  
GN NE1570.  
OS Nitrosomonas europaea.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
OC Nitrosomonadaceae; Nitrosomonas.  
OC NCBI\_TaxID=915;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19718 / IPO 14298;  
RX MEDLINE=22586410; PubMed=12700255;  
RA Chain P., Iamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,  
RA Hauser F., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
RA Arciero D.M., Holmes N.G., Whitaker M.M., Arp D.J.;  
RT "Complete genome sequence of the ammonia-oxidizing bacterium and  
RT obligate chemolithoautotroph Nitrosomonas europaea.";  
RJ J. Bacteriol. 185:2759-2773 (2003).  
DR EMBL: BX31861; CAB5481.1;  
DR GO: GO:0016829; F:lyase activity; IEA.  
DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro: IPR006140; Antifreeze\_dom.  
DR InterPro: IPR006190; Antifreeze\_like.  
DR InterPro: IPR006444; CBS\_domain.  
DR InterPro: IPR004144; Neut.  
DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF00571; CBS; 1.  
DR Pfam: PF03102; Neut; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
SQ SEQUENCE 749 AA; 83956 MW; 284B8128108149A9 CRC64;  
Query Match: 27.9%; Score 526; DB 16; Length 749;  
Best Local Similarity 35.6%; Pred. No. 1.7e-34;  
Matches 115; Conservative 63; Mismatches 139; Indels 6; Gaps 5;

QY 13 VGGGHCFTTIAETGONHGDLDVAKEMIRAKCEGADCAKFOKSELEFFKRNKALERP72  
 DB 125 IADONPAFTTAEVGNHNGDITGLAKELVLAVERAGDCKVFORRDSLSYNSG--RAME 182  
 QY 73 SKHSMWKYV-EKRRHLEPSHDQRELORYAEVGIFFTAAGMDEKAVEFLHELVNPFK 131  
 DB 183 AGVDLGSQVTLDLINKPOLNHDELCOVFYPCQODILPCTPMDVSAVHLDEYGLBAEK 242  
 QY 132 VGGGDTNPFYLEKTKAKGPMWISSGMSMDTMKQVQVYKRLNPFCLQCTSAVPLQ 191  
 DB 243 VASADTNTNEMETLTKATGPPLICSTMSSEADIKGSVDLRLGAPFALLHGNSTYPAE 302  
 QY 192 PEVUNIRVISEYOKLFPDIPIGYSGHETGIAISVAVALGAKYLEHITLIDKTKWGS 251  
 DB 303 FKVNNLNYLPHLMQGLSTV-VGYSGRHGRSPVLAVALGARIVERGFYDRSMGNDHX 361  
 QY 252 ASLEPGEHLAVRSVRLVERALG-SPTKQLPCEMACNEGLGSSVYAKKIPBGTILMD 310  
 DB 362 VSLPPEFAEMVQRINIEBALQGGGRSLTQSEMINTLAKSLVINCDLSQSLIRDS 421  
 QY 311 MLTVKGEPRKAYPPEDITFNLVGK 333  
 DB 422 MITVK-SPGGGLQPNRIDELAGK 443

## RESULT 12

082HY4 PRELIMINARY; PRT; 311 AA.  
 AC 082HY4  
 DT 01-JUN-2003 (TEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Putative N-acetylneuraminic acid (Neu5Ac) synthase.  
 GN SAV3373.  
 OS Streptomyces avermitilis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycetaceae; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxID=33903;  
 RX NCBI\_TaxID=33903;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=21477403; PubMed=11572948;  
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
 RA Kikuchi H., Shida T., Sakaki Y., Hattori M.;  
 RT "Genome sequence of an industrial microorganism Streptomyces  
 RT avermitilis: deducing the ability of producing secondary  
 RT metabolites";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RL (2)  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=22608306; PubMed=11692562;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shida T.,  
 RA Sakaki Y., Hattori M., Omura S.;  
 RT "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis";  
 RT Nat. Biotechnol. 21:526-531(2003).  
 RL  
 RN EMBL, AP005034; BAC71085.1;  
 RC GO:GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO:GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR GO:GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000414; NEUB.  
 DR InterPro: IPR000169; SHprot\_acsite.  
 DR Pfam: PF03102; NEUB.1.  
 DR PROSITE: PS00639; THOL\_PROTEASE\_HIS.1.  
 DR Complete proteome.  
 SQ SEQUENCE 311 AA; 34162 MW; E1B0AFB7CAAF400 CRC64;

Query Match 27.8%; Score 524; DB 16; Length 311;  
 Best Local Similarity 38.9%; Pred. No. 7e-35;  
 Matches 122; Conservative 53; Mismatches 115; Indels 24; Gaps 7;

QY 14 GGHCFPIIAEIGONHGDLDVAKEMIRAKCEGADCAKFOKSELEFFKRNKALERP71  
 DB 15 GGHVYVYVIGLGNHNELENAFLIDAAAAGCDVAKFQRTPEICTPQOMDIR-- 72  
 QY 72 TSXHSWKV-TYEHKRLHLEPSHDQRELORYAEVGIFFTAAGMDEKAVEFLHELVNPFK 130  
 DB 73 --DTWGRNTYIDYHRRVPEGDEYRQIDEVAKSNITOMFASPMTEAVALKEKTDIPAH 130  
 QY 131 KVGSGDTNPFYLEKTKAKGPMWISSGMSMDTMKQVQVYKRLNPFCLQCTSAVPLQ 189  
 DB 131 KVASLNDDELRLALRGKRVIIISTG--ASTPKQLRHAVEVIGSDVILLCHASTVP 187  
 QY 190 LQPEDVNRVISEYOKLFPDIPIGYSGHETGIAISVAVALGAKYLEHITLIDKTKWGS 249  
 DB 188 AKAEEINLRVINTLQAEYVNVPIGSGHETGIAISVAVALGATVEHITLIDRAMGSD 247  
 QY 250 HSASLEPGEHLAVRSVRLVERALGSPTRQLPCEMACNEKLGK--SVYAKKIPBGTIL 307  
 DB 248 QASVPEPQSLTVLVDIRITELSLDGVKQVYSELGPKKRLRRVGVAAEAI----- 301  
 QY 308 TMDMLTVKGEPRKAYPPEDITFNLVGK 321  
 DB 302 -----AAAGEPVA 310

## RESULT 13

07X523 PRELIMINARY; PRT; 334 AA.  
 AC 07X523  
 DT 01-OCT-2003 (TEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Pmc.  
 GN Pmc.  
 OS Campylobacter coli.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 OC NCBI\_TaxID=195;  
 RX NCBI\_TaxID=195;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC167;  
 RX MEDLINE=96423180; PubMed=8825781;  
 RA Querry P., Doig P., Alm R.A., Burr D.H., Kinneela N., Trust T.J.;  
 RT "Identification and characterization of genes required for post-  
 RT translational modification of Campylobacter coli VC167 flagellin";  
 RT Mol. Microbiol. 19:365-378(1996).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=VC167;  
 RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Querry P.;  
 RT "Structural heterogeneity of carbohydrate modifications affects  
 RT serospecificity of Campylobacter flagellins";  
 RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=VC167;  
 RA Querry P.M., Doig P., Alm R.A., Burr D.H., Kinneela N., Trust T.J.;  
 RT Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=VC167;  
 RA Querry P., Ewing C.P., Moran A.P., Trust T.J.;  
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=VC167;  
 RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Querry P.;  
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AY102621; AAM76277.1;  
 SQ SEQUENCE 334 AA; 36905 MW; C415F8473D91D1B6 CRC64;

Query Match 27.5%; Score 518; DB 2; Length 334;  
 Best Local Similarity 36.6%; Pred. No. 2.4e-34;

Matches 123; Conservative 62; Mismatches 135; Indels 16; Gaps 5;

QY 21 IIAETGQNHQGLDVAKRMIRAKKCGADCAKQSELEFKKRLKRPYTSK-HSMGK 79  
 DB 6 IIAAGVNNHGDNLAKKLIEVAAGSADPFAKQSFALCVSKAKKAAVQLTKTADE 65  
 QY 80 TYGKHLRLEFSDQRELORYAEVGIFFTSAGNDMAVEFLHNLNVPFFKVGSGDNTN 139  
 DB 66 SOLKMKLELDPMNAKQLLISHAKOGGIAFLSTADPLESIEHLDLUGLEVFIPSGEITN 125  
 QY 140 FPLYLKTAKKGRPMVIVSSGMSMDTKQYQYV--KPLNPFCHLOCSAIPLOPEBYN 196  
 DB 126 LFLYKKAIAKANKKILISTGMSNIGIEALEVCEKGTORANITLHCTTEYPADFNEN 185  
 QY 197 LFAVISEYOKLPDIPFGYSGHETGIAISVAVALGAKVLERHITLTKTKSGDSHSLER 256  
 DB 186 LKAMOTLKNAF-NLDVGYSDHTKGIHISLIALAGSAVLEKFTLIDKMSGPDHKSLEP 244  
 QY 257 GBIAELVRSVRLVERALGSPYKQLLPCEMAKNEKLGKSVAAKVIPEGTILTMMLTVKV 316  
 DB 245 DEIQELCTKIREIESALGDGIXKASKSEKXNIEIAKKSIVAKKIKTKEGIFSEENLITK- 303  
 QY 317 GEPKAPPEDIPNL-----VGKVLVTVEDDTIME 347  
 DB 304 -----RPASGISAMRYDEYLGKAKASDYEDDELIHE 334

RESULT 14  
 Q8DD27 PRELIMINARY; PRT: 333 AA.

AC Q8DD27; 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Sialic acid synthase.  
 GN VVI0808.  
 OS *Vibrio vulnificus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=672;  
 RN R01111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCPS;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CMCPS."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB016799; AAC09312.1; -  
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro: IPR006014; Antifreeze dom.  
 DR InterPro: IPR006190; Antifreeze\_like.  
 DR Pfam: PF004144; Neut.  
 DR Pfam: PF01354; Antifreeze; 1.  
 DR Pfam: PF03102; Neut; 1.  
 DR PROSITE: PS50844; AFP\_LIKE; 1.  
 DR Complete proteome.  
 KM  
 SQ SEQUENCE 333 AA; 36504 MW; 419D1FLA07EBB2A4 CRC64;

Query Match 27.4%; Score 516.5; DB 16; Length 333;  
 Best Local Similarity 39.0%; Pred. No. 3.2e-34;  
 Matches 117; Conservative 53; Mismatches 125; Indels 5; Gaps 3;

QY 20 FIIAETGQNHQGLDVAKRMIRAKKCGADCAKQSELEFKKRLKRPYTSKSM-G-78  
 DB 6 FIIAAGVNNHGDNLAKKLIDVAAGVAVKQFTWKTELLVTPAKAAVEVENTORE 65  
 QY 79 KTYGKHLRLEFSDQRELORYAEVGIFFTSAGNDMAVEFLHNLNVPFFKVGSGDNTN 138  
 DB 66 EYGFEMKRLLEISYDDETKSYCDGKIFTMSTPDEOSATFLDGLQ-AVFISGELT 124  
 QY 139 NPFYLEKTAKKGRPMVIVSSGMSMDTKQYQYV--KPLNPFCHLOCSAIPLOPEBYN 195  
 DB 125 NTFLEHIAAPAKRVILSTGKGLSEVERAVVTLRDAGSLDMLTVLHATTDPTAFEDV 184

QY 196 NLVYISEYOKLPDIPFGYSGHETGIAISVAVALGAKVLERHITLTKTKSGDSHSLER 255  
 DB 185 NLAMKTEQAFPDIPFGYSDHTLGTETLPAVALGAKVLEKFTLIDKMSGPDHKSLE 244  
 QY 256 GBIAELVRSVRLVERALGSPYKQLLPCEMAKNEKLGKSVAAKVIPEGTILTMMLTVK 315  
 DB 245 DEIQELCTKIREIESALGDGIXKASKSEKXNIEIAKKSIVAKKIKTKEGIFSEENLITK- 304

RESULT 15  
 Q8KNA2 PRELIMINARY; PRT: 361 AA.

AC Q8KNA2; 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Similar to Neut. family.  
 OS *Pseudomonas aeruginosa*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; *Pseudomonas*.  
 OX NCBI\_TaxID=287;  
 RN R01111  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22053227; PubMed=12057956;  
 RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavlin T.V.,  
 RA Ivey R.G., Zhou Y., Kaul R., Ciendoming J.B., Olson M.V.;  
 RT "Genetic Variation at the O-Antigen Biosynthetic Locus in *Pseudomonas aeruginosa*."  
 RT J. Bacteriol. 184:3614-3622(2002).  
 DR EMBL: AF498403; AA027589.1; -  
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro: IPR006014; Antifreeze dom.  
 DR InterPro: IPR006190; Antifreeze\_like.  
 DR Pfam: PF01354; Antifreeze; 1.  
 DR Pfam: PF03102; Neut; 1.  
 DR PROSITE: PS50844; AFP\_LIKE; 1.  
 DR Complete proteome.  
 KM  
 SQ SEQUENCE 361 AA; 39425 MW; E63D3B9BF5DD76D CRC64;

Query Match 27.2%; Score 514; DB 2; Length 361;  
 Best Local Similarity 35.8%; Pred. No. 5.7e-34;  
 Matches 122; Conservative 55; Mismatches 132; Indels 32; Gaps 5;

QY 20 FIIAETGQNHQGLDVAKRMIRAKKCGADCAKQSELEFKKRLKRPYTSKSMGK 79  
 DB 10 FIIAAGVNNHGDNLAKKLIDVAAGVAVKQFTPAKSLASAPKANY-QKHTTDV 68  
 QY 80 TYGE-HKHLRLEFSDQRELORYAEVGIFFTSAGNDMAVEFLHNLNVPFFKVGSGDT 137  
 DB 69 TESQALMKKLELPKEMFELQAHNHGIEFFISTAFSDSLAFLAEKQLEFFKVPSSGL 128  
 QY 138 NPFYLEKTAKKGRPMVIVSSGMSMDTKQYQYV----- 172  
 DB 129 TNGPLLMFAFKTGKRLISTGKATISVEQGLAIYAHLSGDNBPKMDDEVRLKSNPSV 188  
 QY 173 -KPLNPFCHLOCSAIPLOPEBYNLRVISEYOKLPDIPFGYSGHETGIAISVAVALG 231  
 DB 189 RMQLOGHVSLHCTSYQYPTPPDEVNLLMDTLRSF--GLAVGSDHTGGLVPLAAVARG 246  
 QY 232 AKVLERHITLTKTKSGDSHSLERGBIAELVRSVRLVERALGSPYKQLLPCEMAKNEK 291  
 DB 247 ACTIKKHTTLDRSMGPHKASLERGELAAGVAAQIRMLEVALGSPYKAPQPSBEMTRQAA 306  
 QY 292 GKSVAAYKIPBGTILTMMLTVKVEBPRAVPEPDIPNLVG 332  
 DB 307 KQVVAARDIEAGMIITRDILT-ARSGHGLPPTSLMELVG 346

Search completed: May 6, 2004, 09:07:35  
 Job time: 37.3895 secs

Thu May 13 11:53:08 2004

us-09-930-440b-6.ra1

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 6, 2004, 08:59:54 ; Search time 14.9474 Seconds  
(without alignments)  
1239.925 Million cell updates/sec

Title: US-09-930-440B-6

Perfect score: 1887  
Sequence: 1 MPELELCGRWVGQHPHF.....EEDDTMEELVDNHKKIKS 359

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: 1: Issued patents AA:\*  
2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1887	100.0	US-09-516-143A-4	Sequence 4, Appl1
2	336.5	17.8	US-09-455-406-21	Sequence 21, Appl1
3	113.5	6.0	US-09-107-532A-6711	Sequence 6711, Ap
4	100	5.3	US-09-648-004-10	Sequence 10, Appl1
5	93.5	5.0	US-09-134-000C-5846	Sequence 5846, Ap
6	93	4.9	US-09-252-991A-24216	Sequence 24216, A
7	93	4.9	US-08-907-166-2	Sequence 2, Appl1
8	93	4.8	US-09-391-340-2	Sequence 2, Appl1
9	92.5	4.9	US-09-343-681A-5022	Sequence 5022, Ap
10	92	4.9	US-09-498-520A-12	Sequence 12, Appl1
11	91.5	4.8	US-09-252-991A-25083	Sequence 25083, A
12	91	4.8	US-09-134-001C-3575	Sequence 3575, Ap
13	90	4.8	US-09-134-001C-4820	Sequence 4820, Ap
14	89	4.7	US-09-269-861A-8	Sequence 8, Appl1
15	88.5	4.7	US-09-134-000C-5157	Sequence 5157, Ap
16	88.5	4.7	US-09-337-913-1	Sequence 1, Appl1
17	88.5	4.7	US-08-750-524-1	Sequence 1, Appl1
18	88	4.7	US-08-492-027A-8	Sequence 15, Appl1
19	87	4.6	US-08-913-942-15	Sequence 26, Appl1
20	87	4.6	US-09-268-347-26	Sequence 6427, Ap
21	87	4.6	US-09-134-000C-6427	Sequence 16, Appl1
22	86.5	4.6	US-08-347-801-16	Sequence 142, Appl1
23	86.5	4.6	US-08-311-731A-142	Sequence 11591, A
24	86	4.6	US-09-489-039A-11591	Sequence 3, Appl1
25	85.5	4.5	US-07-768-286B-6	Sequence 3, Appl1
26	85.5	4.5	US-08-487-823B-3	Sequence 3, Appl1
27	85.5	4.5	US-08-997-040-3	Sequence 3, Appl1

28	85.5	4.5	382	2	US-09-203-237-3	Sequence 3, Appl1
29	85.5	4.5	760	3	US-08-928-941D-29	Sequence 29, Appl1
30	85.5	4.5	760	4	US-09-280-590A-29	Sequence 29, Appl1
31	85.5	4.5	760	4	US-09-892-398-29	Sequence 29, Appl1
32	85	4.5	789	4	US-09-134-000C-4939	Sequence 4939, Ap
33	85	4.5	1201	4	US-09-328-352-5890	Sequence 5890, Ap
34	84.5	4.5	303	4	US-08-928-941D-18	Sequence 18, Appl1
35	84.5	4.5	303	4	US-09-280-590A-18	Sequence 18, Appl1
36	84.5	4.5	303	4	US-09-892-398-18	Sequence 18, Appl1
37	84.5	4.5	323	2	US-09-019-216-3	Sequence 3, Appl1
38	84.5	4.5	392	4	US-09-249-241-3	Sequence 4, Appl1
39	84.5	4.5	392	1	US-07-768-286B-4	Sequence 7, Appl1
40	84.5	4.5	405	1	US-08-121-714-7	Sequence 7, Appl1
41	84.5	4.5	405	1	US-08-477-112-7	Sequence 7, Appl1
42	84.5	4.5	405	2	US-08-477-112-7	Sequence 7, Appl1
43	84.5	4.5	405	5	PCT-US93-08322-7	Sequence 7, Appl1
44	84.5	4.5	415	1	US-07-911-531-19	Sequence 19, Appl1
45	84.5	4.5	415	1	US-07-693-636A-19	Sequence 19, Appl1

## ALIGNMENTS

RESULT 1  
US-09-516-143A-4  
Sequence 4, Application US/09516143A  
Patent No. 633182  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Human Glycosylation Enzymes  
FILE REFERENCE: PF0505PCT  
CURRENT APPLICATION NUMBER: US/09/516,143A  
PRIORITY FILING DATE: 2000-03-01  
PRIORITY FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 359  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-516-143A-4

Query Match 100.0%; Score 1887; DB 4; Length 359;  
Best Local Similarity 100.0%; Pred. No. 2.8e-193;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MPELELCGRWVGQHPHFIIAETGONHGGDLDVAKRMIRAKKGGADCAKFKSELEF	60
DB	1	MPELELCGRWVGQHPFIIAETGONHGGDLDVAKRMIRAKKGGADCAKFKSELEF	60
QY	61	KFNKKALERPPTYSKSHGWKTYGKHKLHFSHDYRELCRYAEVGFPTASGDMAYE	120
DB	61	KFNKKALERPPTYSKSHGWKTYGKHKLHFSHDYRELCRYAEVGFPTASGDMAYE	120
QY	121	FLHEINVPFFKSGSTNNPPIYKAKKGRPVVTSQMSMDTKOVQIVELPNFC	180
DB	121	FLHEINVPFFKSGSTNNPPIYKAKKGRPVVTSQMSMDTKOVQIVELPNFC	180
QY	181	FLQCTSAVPLQPEEDVLAIVISEYOKLPDIPIGYSGHETGIALSVAAVAGAVLERHIT	240
DB	181	FLQCTSAVPLQPEEDVLAIVISEYOKLPDIPIGYSGHETGIALSVAAVAGAVLERHIT	240
QY	241	LDTYWGSHSASLEGEALAEVRSVLRALAGSTTKQLLPBMAENELGKSVYAKK	300
DB	241	LDTYWGSHSASLEGEALAEVRSVLRALAGSTTKQLLPBMAENELGKSVYAKK	300
QY	301	IPGGTILTMKTLVKGKPRKAPPEDEIENLVGKQVLTVEEDDTIMEELVDNHKKIKS	359
DB	301	IPGGTILTMKTLVKGKPRKAPPEDEIENLVGKQVLTVEEDDTIMEELVDNHKKIKS	359

RESULT 2

```

US-09-495-406-21
; Sequence 21, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; TITLE OF INVENTION: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 21
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: static acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of LOS biosynthesis locus)
US-09-495-406-21

Query Match      17.8%; Score 336.5; DB 4; Length 346;
Best Local Similarity 29.1%; Pred. No. 2,4e-27;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHGGDIDVAKRMIRAKGCGADCAFKQSELEFKRRLALR-REYTSKHSWGK 79
DB 19 VPEIGINNGSLHAKIVDAAFSTGAKIIGHQHIHVEDMSKXAKVTPENAKIS--- 75
QY 80 TYGEKRLHLESHDQRELOIAEYVGIFFTSAGDMNAVELEHINLPFVSGDITNN 139
DB 76 -IYEIMQKCALDYKELALKEYTEKLGIVYSTPPSRACANLEDMGVSAFQIGSGECNN 134
QY 140 FYLEETAKKGRPVYSSGMSQMDTMKQVOIVKPLNPFCLQCTSAVPLQPEBVTNRV 199
DB 135 YPLIKIILAFKKEIVYISGMSISIKPTVKILLDNHIFVLEMTNLYPTPHNIVRLNA 194
QY 200 ISEYQCLFPDIPIGVSGHETGIAISVAVALGAKVLERHITLDKTKSGSDHSASLEPSEL 239
DB 195 MELKEKF-SCWGLSDHTDNLACIGAVAGACVLERHFTDSMRSGPDIVCSMDTQAL 253
QY 260 AEL-VASVALVERALASPTKOLLPCMAENKLGKSVAKVKIPRSTLITMDMLTVKVG 318
DB 254 KELTIQSEQWAINRGNNESKKAQKQVITIDFAPASVSIKDKKGEVLSMDNINVKKPG 313
QY 319 PRAVPEPIDFNLVYKGLVTVEDDTIMEE 348
DB 314 LGGISAEPENILGKALRDIENTDLSYE 343

RESULT 3
US-09-107-532a-6711
; Sequence 6711, Application US/09107532a
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC

```

```

OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532a
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6711:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...343
SEQUENCE DESCRIPTION: SEQ ID NO: 6711:
US-09-107-532a-6711

Query Match      6.0%; Score 113.5; DB 4; Length 343;
Best Local Similarity 22.4%; Pred. No. 0.0017;
Matches 65; Conservative 48; Mismatches 102; Indels 75; Gaps 4;

QY 10 GRWVGQHPCEITAEIGNHGGDIDVAKRMIRAKGCGADCAFKQSELEFKRRLALR 63
DB 96 GSFYVAGPCEI-----EG-LDOIRECARAKAKAGA-----KILGAGFK 134
QY 70 PYTSKHSNGKTYGSHKHLESHDQRELOIAEYVGIFFTSAGDMNAVELEHINLPF 129
DB 135 PRISYPA--QGLESGDKT-----IROADRPKAKVITETMDGHIDVAEYS-DI 183
QY 130 FKVSGDITNNPYLEKTAKKRPNVYSSGMSQMDTMKQ---VYQIVKPLNPFCLQ- 183
DB 184 LQIGARNQNKILSAVCKTGKPGVGLKRGISG--TINEMINAAEYIAVEDKSP-VIFER 240
QY 184 -----CTSAVPLQPEBVTNRVISEYQCLFPDIPIGVSGHETG-----AISV 225
DB 241 GIRYETATNTFPLSAVPLKGLTHRPVIVD-----PSHOTGELWELVPPRAR 288
QY 226 AAVALGA--KVLERHITLDKTKSGSDHSASLEPSELAEVRSVLEVERAL 273
DB 289 AGVASGADGMIVELHPPDANAM--SDGPQSLNEXKTYRGMKEVAIMKEAM 336

RESULT 4
US-09-648-004-10
; Sequence 10, Application US/09648004
; Patent No. 6498242
; GENERAL INFORMATION:
; APPLICANT: CHEN, OIONG
; APPLICANT: THOMAS, STUART
; TITLE OF INVENTION: BIOMEDICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; TITLE OF INVENTION: INTERMEDIATES
; FILE REFERENCE: CL-1341-A
; CURRENT APPLICATION NUMBER: US/09/648,004
; CURRENT FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/252,553
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32

```

SOFTWARE: Microsoft Office 97  
 SEQ ID NO 10  
 LENGTH: 547  
 TYPE: PRF  
 ORGANISM: Acinetobacter sp.  
 US-09-648-004-10

Query Match  
 Best Local Similarity 20.5%; Score 100; DB 4; Length 547;  
 Matches 82; Conservative 65; Mismatches 119; Indels 134; Gaps 24;

QY 4 ELET-----CPGRVVGQHP-----CFI-IAEIQGNHGGDLDAKRMIR 41  
 DB 32 ELETQADKADKADVAGTWTNNRYPGALSDTEHLKYCSWDKELLOS---LEIKKKYQ 87  
 QY 42 MAKECGADCAKFOKSELEFPKPKKALERPYSKSHNGKTYGKHRLLEFSDQYRE--L 98  
 DB 88 -----GPDVAKYIQVAA-----KHDIKSY-QFTAVQSAH--YNEADAL 125  
 QY 99 QYAAEVGIFETASGMDMAVEBELHNLNP-----FFKVGSGDENPNPYLEKTAKKGRP 152  
 DB 126 WEVTTYGDKTKRFL-ITRLGLSLAPNLPNTKGINQFKEHLHTSRWP--DDVSTEGGR 182  
 QY 153 M-VISSGQSMQMDTKQVQVQVKEPLNPFELQCTSAV-----PLQPEDVNLKRVISEQ 204  
 DB 183 VGVIGTG---STGVQVITAVAPLACHLTVFGSAQYSPVIGNDPLSEBDVK-KIKNDYD 237  
 QY 205 KLFPPD---PIGSGHETGIAIVAAVAGAKYLEHITLDKTKW-----246  
 DB 238 KMDGVNAGSALARGNES---TPAMSVSAE--EKKAVFEKXMQTGGGFRFMEFTEDI 291  
 QY 247 GSDHSASLE-----PSELALVSVRLVERALGSPKQKLLPCSMACNEKLKGS-----294  
 DB 292 ATMEANIEAQNIRKIRAEIVDDPALAK-----LMPQDLVAKRPLCSGGYNTFN 343  
 QY 295 -----VYAK-----VKIPGTTITMDMLTVKVG 317  
 DB 344 RDNVRLDEVKAMPIVEITENGVLKNGDFVELDMLICATG 383

RESULT 5  
 US-09-134-000C-5846  
 Sequence 5846, Application US/09134000C  
 Patent No. 6617156  
 GENERAL INFORMATION:  
 APPLICANT: Lynn Doucette-Stamm et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 032796-032  
 CURRENT APPLICATION NUMBER: US/09/134,000C  
 CURRENT FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/055,778  
 PRIOR FILING DATE: 1997-08-15  
 NUMBER OF SEQ ID NOS: 6812  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 5846  
 LENGTH: 389  
 TYPE: PRF  
 ORGANISM: Enterococcus faecalis  
 US-09-134-000C-5846

Query Match  
 Best Local Similarity 20.0%; Score 93.5; DB 4; Length 389;  
 Matches 73; Conservative 52; Mismatches 119; Indels 121; Gaps 17;

QY 38 RMIRAKKCGADCAKFOKSELE-----EFKPKALERPYSK-----HSM 77  
 DB 6 RMIIQKAVMIN-EGQPAQVEVSEORIVAILPEYNEKAVDKYGEKRILPGFIDHDM 64  
 QY 78 GKTYGKHRLLEFSHOYRELQYAEVGIFFASGMDMAVEBELHNLNPFFKVG-----134  
 DB 65 --HGADANHA--NHEPIKEMQAYLPEBGI---TAFLEPTTSTTFKDLHSHSEFVIGSFL 116

QY 135 --GOTTNPFYLEKTAKKGRPVISS-----GMSQMDTKQVQVQVPLNPFEL 182  
 DB 117 BDQGTNGAQIIGIIRH--GPMISSEFFGSHNPBELLVKSVEFTFKMOELAK-----165  
 QY 183 OCTSAVPLQPEDVNLKRVISEYOKLPPDIPIGVSGHETGIAISV-----AAVALG 231  
 DB 166 GHIKMLTAPENDVENALTTV-----CHEHDVVISIGHTAATVEQAMAAVEAG 213  
 QY 222 AKVLERHITLDKTKGSDHSASLEBEL-----AELVRSVRLVERALGSGTKQL 281  
 DB 214 AK-----SFHTFTGEMDISHRKPTAVAAALDSEETFAELIADGVHADYSL-----259  
 QY 282 PCMACNEKLKGSVAAKVIPEGTITMDMLTVKGSBKAAP-PEDIFNLVGRKYLATVE 340  
 DB 260 -----VRVLAKLKGMDYLAVTDSIAKGCQGVYPKPE-----KGIEMVTD 301  
 QY 341 EDDTI 345  
 DB 302 EQNVV 306

RESULT 6  
 US-09-252-991A-24216  
 Sequence 24216, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 10/196,136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,768  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,160  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 24216  
 LENGTH: 375  
 TYPE: PRF  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-24216

Query Match  
 Best Local Similarity 22.5%; Score 93; DB 4; Length 375;  
 Matches 45; Conservative 32; Mismatches 67; Indels 56; Gaps 9;

QY 81 YGEHRLLEFSH-----DOYRELQYAEVGIFFASGMDMAVEBELHNLNPFFKY 132  
 DB 162 YFDGSDHDLNDEYVPRRFSBHYGATKYQAEGLVASADLGEVLALR-----PRFV 214  
 QY 133 GSGDTNPNPYLEKTAKKGRPVISSGMSQMDTKQVQVQVPLNPFELQCTSAVPLQ 192  
 DB 215 GAGDTISIEPRMIOARKRRLILANGLRVD-----FTSVENLNDALFSLIAG--BP 265  
 QY 193 EDVNLKRVISEYOKL-FPDI-----PIGSGH-ETGIAISVAAV-----228  
 DB 266 ALGRYTNISNGQPVFPWADVNVVWQDLDPVG--GHLPYANGYGLAALNEGVCILPGR 323  
 QY 229 -----ALGAKVLERHITLD 242  
 DB 324 PEPVLFRLGMAYMAKNTFTD 343

RESULT 7  
 US-08-907-166-2  
 Sequence 2, Application US/08907166  
 Patent No. 5948666  
 GENERAL INFORMATION:  
 APPLICANT: Callen, Walter  
 TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES  
 FILE REFERENCE: 09010/027001

Thu May 13 11:53:08 2004

us-09-930-440b-6.rai

Page 4

CURRENT APPLICATION NUMBER: US/08/907.166  
CURRENT FILING DATE: 1997-08-06  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 868  
TYPE: PRT  
ORGANISM: Ammonifex degensis  
US-08-907-166-2

Query Match 4.9%; Score 93; DB 2; Length 868;  
Best Local Similarity 19.8%; Pred. No. 1.2;  
Matches 72; Conservative 55; Mismatches 102; Indels 134; Gaps 17;

51 AKFQSELEFFKFNKALRPTYSKISWGTGKHEHLEFSDQYRELQRYAEVGIFFT 110  
57 AAFDKSKITF-----RHAALETY---KAHRPATPDELROQFNLIKGV----- 95  
111 ASGMDMAVEFLHEINVPFFKVGSGDTNNF--PYLEKTAKKGRPMVVISGMSMDTKQV 168  
96 -----LTALNVFVEKEGFEADDLIGTLVRAEKGWOCCLIVTG--DIDLAL--- 139  
169 YQIVKPLNPFCLQCTSAVPLQPE-----DYNLRVISEYOKL-----PPDIPIGY 214  
140 -QIVSPITTVLMKRGISEIAVFNBAEYKRRFGVTPRQLPDKLAGDASDNIPGLP--- 195  
215 SGHETGIAISVAVALGAKVLERHTITLDTKWSGDSASLPEGLAEIVRSVLRVERALG 274  
196 -----GIGPKYA-----SRLLQSHQSLKLE---SKEFFPAKLEETLE--RHKEEAVL 239  
225 SPTKQLPCMACNENK-----GKSVAAKVI-----PEGTILITMDML 312  
240 AKGLALIRVDVPLEEETIRPWPFPNLTLEVFSTRLEFRTAKRLELFPFARLLSASGL 299  
313 T-----VKNGEPR-----AYPP-----EDIFNLVVK 333  
300 TPSAVRVKVERPRELERLGELGROFPALAVPVLRRKATSSFLALCLGGEKVFLEEGP 359  
334 KVL 336  
360 EVL 362

RESULT 8  
US-09-391-340-2  
Sequence 2, Application US/09391340A  
Patent No. 6492511  
GENERAL INFORMATION:  
APPLICANT: Callen, Walter  
APPLICANT: Mather, Eric  
TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES  
FILE REFERENCE: 09010/027001  
CURRENT APPLICATION NUMBER: US/09/391,340A  
CURRENT FILING DATE: 1999-09-07  
EARLIER APPLICATION NUMBER: US 08/907,166  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 868  
TYPE: PRT  
ORGANISM: Ammonifex degensis  
US-09-391-340-2

Query Match 4.9%; Score 93; DB 4; Length 868;  
Best Local Similarity 19.8%; Pred. No. 1.2;  
Matches 72; Conservative 55; Mismatches 102; Indels 134; Gaps 17;

51 AKFQSELEFFKFNKALRPTYSKISWGTGKHEHLEFSDQYRELQRYAEVGIFFT 110  
57 AAFDKSKITF-----RHAALETY---KAHRPATPDELROQFNLIKGV----- 95  
111 ASGMDMAVEFLHEINVPFFKVGSGDTNNF--PYLEKTAKKGRPMVVISGMSMDTKQV 168

96 -----LTALNVFVEKEGFEADDLIGTLVRAEKGWOCCLIVTG--DIDLAL--- 139  
169 YQIVKPLNPFCLQCTSAVPLQPE-----DYNLRVISEYOKL-----PPDIPIGY 214  
140 -QIVSPITTVLMKRGISEIAVFNBAEYKRRFGVTPRQLPDKLAGDASDNIPGLP--- 195  
215 SGHETGIAISVAVALGAKVLERHTITLDTKWSGDSASLPEGLAEIVRSVLRVERALG 274  
196 -----GIGPKYA-----SRLLQSHQSLKLE---SKEFFPAKLEETLE--RHKEEAVL 239  
275 SPTKQLPCMACNENK-----GKSVAAKVI-----PEGTILITMDML 312  
240 AKGLALIRVDVPLEEETIRPWPFPNLTLEVFSTRLEFRTAKRLELFPFARLLSASGL 299  
313 T-----VKNGEPR-----AYPP-----EDIFNLVVK 333  
300 TPSAVRVKVERPRELERLGELGROFPALAVPVLRRKATSSFLALCLGGEKVFLEEGP 359  
334 KVL 336  
360 EVL 362

RESULT 9  
US-09-543-681A-5022  
Sequence 5022, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI  
FILE REFERENCE: 2709, 1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 5022  
LENGTH: 753  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-5022

Query Match 4.9%; Score 92.5; DB 4; Length 753;  
Best Local Similarity 20.8%; Pred. No. 1.1;  
Matches 69; Conservative 58; Mismatches 105; Indels 99; Gaps 18;

117 MAVEFLHEINVPFFKVGSGDTNNFPY-----LEKTAKKGRPMVVISGMS 161  
152 IAKDPLDINFPVLYDQKTYTLPIYGEOSMEKKRTLELEKLABEKAAYQYLEN 211  
162 MDTKQ-----VYQIVKPLNPFCLQCTSAVPLQPEEDYNLRVISEY----- 203  
212 KDSVQKQKTKMIRPIEEKDPOMWYK-----YKLISQDINILFDSAYLVNPPATYILY 267  
204 -----GKLEPP-----DIPICYS---GHEGIAISVAALV---ALGA---KYLERH 238  
268 QIVKGDKNVPLQNSFLANILYIKKSDLAHBOGLIKIEILNNSNTLSSYFNNLEKY 327  
239 I-TLDTKWSGDH-----SASLEP--GELAEIVRSVLRVERALG--SPTKQLPCMAC 287  
328 IDTLLKTPNSPQAMALANVYIKGNSFYNEKALALVBAVNIQSPSKILLALYS 387  
288 N-EKLGKSVAAKVIPEGTILITMDMLT-----VKV-----GEKXATPEPDIPLVNOCKV 335  
388 NSEGVHONIRKAVSELEKENTINDKLTGKQRELVKTYFDGASDYIKKEIVNLTRESV 446  
336 LV-----TVEED-----DTIMEELVDNH 353  
447 IKNSAGTINONISLAFYADLILEEDVANN 477







Thu May 13 11:53:08 2004

us-09-930-440b-6.rai

Page 7

```
QY 137 TNNF-----PYLEKTA---KGRPMV1SSGMSMDTMKQYQIVKPELNPFCF 181
Db 150 NQNFLLTILSHSGEKKYKLTALAKKXKPIISFVGAKN-STLGRLLADLVFSTDsysPF 208
QY 182 LQCTSAVPIQPEVDVNLKRVISYQKLPFDIPIGYSGHFTGIAIS 224
Db 209 ---STVAQPCMFPGQTLITFALI---CAYLNHEDSIPIS 243
```

Search completed: May 6, 2004, 09:08:38  
Job time : 16.9474 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 38.8108 Seconds  
(Without alignments)  
2567.492 Million cell updates/sec

Title: US-09-930-440b-6  
Sequence: 1 MPELELCGRWVGQHPGCF.....EEDDTIMEELVDNKGKINS 359

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PTCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PTCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1887	100.0	359	US-09-984-205-4	Sequence 4, Appl1
2	1887	100.0	359	US-09-930-440b-6	Sequence 6, Appl1
3	1883	99.8	359	US-10-264-237-2597	Sequence 2597, Ap
4	544.5	28.9	338	US-09-767-041-24	Sequence 24, Appl
5	524	27.8	311	US-10-156-761-10910	Sequence 10910, A
6	510.5	27.1	346	US-09-930-440b-8	Sequence 8, Appl1
7	381	20.2	123	US-10-106-698-6926	Sequence 6926, Ap
8	336.5	17.8	346	US-09-816-0284-35	Sequence 35, Appl
9	336.5	17.8	346	US-10-303-161-35	Sequence 35, Appl
10	336.5	17.8	346	US-10-303-118-35	Sequence 35, Appl
11	336.5	17.8	346	US-10-303-128-35	Sequence 35, Appl
12	336.5	17.8	346	US-10-303-134-35	Sequence 35, Appl
13	336.5	17.8	346	US-10-303-162-35	Sequence 35, Appl
14	303	16.1	340	US-09-882-227-490	Sequence 490, Appl
15	287	15.2	340	US-10-335-977-7626	Sequence 7626, Ap

16	283.5	15.0	314	US-10-335-977-7625	Sequence 7625, Ap
17	122.5	6.5	205	US-10-321-396C-4	Sequence 4, Appl1
18	119	6.3	352	US-10-369-493-18790	Sequence 18790, A
19	117.5	6.2	275	US-10-321-396C-2	Sequence 2, Appl1
20	116	6.1	352	US-10-369-493-20092	Sequence 20092, A
21	113.5	6.0	333	US-10-282-122A-57486	Sequence 57486, A
22	113	6.0	285	US-10-369-493-18119	Sequence 18119, A
23	110	5.8	135	US-10-321-396C-6	Sequence 6, Appl1
24	110	5.8	265	US-10-369-493-31568	Sequence 21568, A
25	109	5.8	337	US-10-369-493-9746	Sequence 9746, Ap
26	108.5	5.7	335	US-10-369-493-50181	Sequence 20181, A
27	108.5	5.7	333	US-10-369-493-2631	Sequence 2631, Ap
28	108.5	5.7	633	US-10-282-122A-63506	Sequence 63506, A
29	108	5.7	338	US-10-369-493-2930	Sequence 2930, Ap
30	107.5	5.7	337	US-10-282-122A-51516	Sequence 51516, A
31	107	5.7	276	US-10-369-493-22853	Sequence 22853, A
32	107	5.7	330	US-10-282-122A-67597	Sequence 67597, A
33	103.5	5.5	329	US-10-369-493-13704	Sequence 13704, A
34	101.5	5.4	335	US-10-369-493-8945	Sequence 8945, Ap
35	101.5	5.4	337	US-10-282-122A-53304	Sequence 53304, A
36	101.5	5.4	341	US-10-282-122A-57228	Sequence 57228, A
37	101	5.4	1175	US-10-369-493-5504	Sequence 6504, Ap
38	100	5.3	547	US-10-272-419-10	Sequence 10, Appl
39	100	5.3	547	US-10-282-122A-70920	Sequence 20, Appl
40	100	5.3	1108	US-10-087-192-1206	Sequence 1206, Appl
41	100	5.3	1189	US-10-282-122A-70920	Sequence 70920, A
42	97.5	5.2	640	US-10-369-493-16235	Sequence 16235, A
43	97.5	5.2	641	US-10-369-493-15488	Sequence 15488, A
44	97.5	5.2	641	US-10-369-493-15860	Sequence 15860, A
45	97.5	5.2	642	US-10-389-566-1089	Sequence 1089, Ap

ALIGNMENTS

RESULT 1  
US-09-984-205-4  
Sequence 4, Application US/09984205  
Patent No. US20020137175A1  
GENERAL INFORMATION:  
APPLICANT: Coleman, Timothy A. et al.  
TITLE OF INVENTION: Human Glycosyltransferase  
FILE REFERENCE: PF505D1  
CURRENT APPLICATION NUMBER: US/09/984,205  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: PCT/US00/05325  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/516,143  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/122,409  
PRIOR FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 359  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-984-205-4  
Query Match 100.0%; Score 1887; DB 9; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.5e-182;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPELELCGRWVGQHPGCFIIAIGQNHQGLDVAKRMIRPAKCEGADCAKFOKSELEF 60  
DB 1 MPELELCGRWVGQHPGCFIIAIGQNHQGLDVAKRMIRPAKCEGADCAKFOKSELEF 60  
QY 61 KERRKALEPPTYSKSHGKTYGHEKHLEFSDQVREIOLRYAEVGIPTASGDMAYE 120  
DB 61 KERRKALEPPTYSKSHGKTYGHEKHLEFSDQVREIOLRYAEVGIPTASGDMAYE 120  
QY 121 FLHLELVPPFKVGSQDTNNPYLEKTAKKGRPVYSSGQSHDTMKQVQIYKPLNPNFC 180  
DB 121 FLHLELVPPFKVGSQDTNNPYLEKTAKKGRPVYSSGQSHDTMKQVQIYKPLNPNFC 180

Db 121 FLHELNVFFKVGSGDTNNPPYLEKTAAGKGPWVSSGMSQMDTMKVYQIVKELNPNFC 180  
 Qy 181 FLQCTSAVPLQPEVDNLRVISEYQKLPDPIPIGSGHETGIAISVAANALCAKLEHRT 240  
 Db 181 FLQCTSAVPLQPEVDNLRVISEYQKLPDPIPIGSGHETGIAISVAANALCAKLEHRT 240  
 Qy 241 LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKQLPCEMACNEKLGKSVAAKVK 300  
 Db 241 LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKQLPCEMACNEKLGKSVAAKVK 300  
 Qy 301 IPBGTILTMQMLTVKVGSPKAYPPEDIFNLVGGKVLVTVVEEDDTIMEELVNHGKIKS 359  
 Db 301 IPBGTILTMQMLTVKVGSPKAYPPEDIFNLVGGKVLVTVVEEDDTIMEELVNHGKIKS 359

RESULT 2  
 US-09-930-440b-6  
 ; Sequence 6, Application US/09930440B  
 ; Patent No. US20020142386A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Betenbaugh et al.  
 ; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways  
 ; FILE REFERENCE: P05092  
 ; CURRENT APPLICATION NUMBER: US/09/930,440B  
 ; PRIOR FILING DATE: 2001-08-16  
 ; PRIOR APPLICATION NUMBER: 60/227,579  
 ; PRIOR FILING DATE: 2000-08-25  
 ; PRIOR APPLICATION NUMBER: 09/516,793  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: 60/169,624  
 ; PRIOR FILING DATE: 1999-12-08  
 ; PRIOR APPLICATION NUMBER: 60/122,582  
 ; PRIOR FILING DATE: 1999-03-02  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 359  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-930-440b-6

Query Match 100.0%; Score 1887; DB 9; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-182;  
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MPELELCPRGWVGQHPCEFIIEIGNHOGDDVAKRMIRMAKCGADCAKFOKSELEF 60  
 Db 1 MPELELCPRGWVGQHPCEFIIEIGNHOGDDVAKRMIRMAKCGADCAKFOKSELEF 60  
 Qy 61 KFNKALERPPTSGHSGKTYGKHKHLESHDQYRELORVABVGIFFTASGMDMAVE 120  
 Db 61 KFNKALERPPTSGHSGKTYGKHKHLESHDQYRELORVABVGIFFTASGMDMAVE 120  
 Qy 121 FLHELNVFFKVGSGDTNNPPYLEKTAAGKGPWVSSGMSQMDTMKVYQIVKELNPNFC 180  
 Db 121 FLHELNVFFKVGSGDTNNPPYLEKTAAGKGPWVSSGMSQMDTMKVYQIVKELNPNFC 180  
 Qy 181 FLQCTSAVPLQPEVDNLRVISEYQKLPDPIPIGSGHETGIAISVAANALCAKLEHRT 240  
 Db 181 FLQCTSAVPLQPEVDNLRVISEYQKLPDPIPIGSGHETGIAISVAANALCAKLEHRT 240  
 Qy 241 LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKQLPCEMACNEKLGKSVAAKVK 300  
 Db 241 LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKQLPCEMACNEKLGKSVAAKVK 300  
 Qy 301 IPBGTILTMQMLTVKVGSPKAYPPEDIFNLVGGKVLVTVVEEDDTIMEELVNHGKIKS 359  
 Db 301 IPBGTILTMQMLTVKVGSPKAYPPEDIFNLVGGKVLVTVVEEDDTIMEELVNHGKIKS 359

RESULT 3  
 US-10-264-237-2597  
 ; Sequence 2597, Application US/10264237

; Publication No. US20040009491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birst et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P013121  
 ; CURRENT APPLICATION NUMBER: US/10/264,237  
 ; PRIOR FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450  
 ; PRIOR FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: US 60/205,515  
 ; PRIOR FILING DATE: 2000-05-19  
 ; NUMBER OF SEQ ID NOS: 2876  
 ; SOFTWARE: Patent In Ver. 3.1  
 ; SEQ ID NO 2597  
 ; LENGTH: 359  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-264-237-2597

Query Match 99.8%; Score 1883; DB 15; Length 359;  
 Best Local Similarity 99.7%; Pred. No. 3.7e-182;  
 Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 MPELELCPRGWVGQHPCEFIIEIGNHOGDDVAKRMIRMAKCGADCAKFOKSELEF 60  
 Db 1 MPELELCPRGWVGQHPCEFIIEIGNHOGDDVAKRMIRMAKCGADCAKFOKSELEF 60  
 Qy 61 KFNKALERPPTSGHSGKTYGKHKHLESHDQYRELORVABVGIFFTASGMDMAVE 120  
 Db 61 KFNKALERPPTSGHSGKTYGKHKHLESHDQYRELORVABVGIFFTASGMDMAVE 120  
 Qy 121 FLHELNVFFKVGSGDTNNPPYLEKTAAGKGPWVSSGMSQMDTMKVYQIVKELNPNFC 180  
 Db 121 FLHELNVFFKVGSGDTNNPPYLEKTAAGKGPWVSSGMSQMDTMKVYQIVKELNPNFC 180  
 Qy 181 FLQCTSAVPLQPEVDNLRVISEYQKLPDPIPIGSGHETGIAISVAANALCAKLEHRT 240  
 Db 181 FLQCTSAVPLQPEVDNLRVISEYQKLPDPIPIGSGHETGIAISVAANALCAKLEHRT 240  
 Qy 241 LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKQLPCEMACNEKLGKSVAAKVK 300  
 Db 241 LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKQLPCEMACNEKLGKSVAAKVK 300  
 Qy 301 IPBGTILTMQMLTVKVGSPKAYPPEDIFNLVGGKVLVTVVEEDDTIMEELVNHGKIKS 359  
 Db 301 IPBGTILTMQMLTVKVGSPKAYPPEDIFNLVGGKVLVTVVEEDDTIMEELVNHGKIKS 359

RESULT 4  
 US-09-767-041-24  
 ; Sequence 24, Application US/09767041  
 ; Patent No. US20020055168A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Smith, Hilda  
 ; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS  
 ; FILE REFERENCE: 2183-4726  
 ; CURRENT APPLICATION NUMBER: US/09/767,041  
 ; PRIOR FILING DATE: 2001-01-22  
 ; PRIOR APPLICATION NUMBER: PCT/NL99/00460  
 ; PRIOR FILING DATE: 1999-07-19  
 ; PRIOR APPLICATION NUMBER: EP98202465.5  
 ; PRIOR FILING DATE: 1998-07-22  
 ; PRIOR APPLICATION NUMBER: EP98202467.1  
 ; PRIOR FILING DATE: 1998-07-22  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SOFTWARE: Patent In version 3.0  
 ; SEQ ID NO 24  
 ; LENGTH: 338  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus suis  
 ; FEATURE: NAME/KEY: misc.feature  
 ; OTHER INFORMATION: CP52P

Thu May 13 11:53:08 2004

us-09-930-440b-6.rapb

**Page 3**

US-09-767-041-24

Query Match	28.9%	Score 544.5	DB 9	Length 338
Best Local Similarity	36.0%	Pred. No. 2.7e-46		
Matches 123	Conservative 73	Mismatches 129	Indels 17	Gaps 6

```

QY 20 FIIIEIGNHODLIVARIMTIMARCCADKAFQSEIEEPFKKALKEEPIKSHSWR 79
Db 3 YIIAIEIGNHGDVILARKWAVEAANDCGDVAVFOIEKADILLISKAPAEV-OKITTE 61
QY 80 TYG--EHKSHLEFSDOYREI,ORVA,BEVCI,FTASGMDMAVEFLHVLNVPFFKVSQDT 137
Db 62 SDSOLEMTRRELTSFSTYDLDRLD,CI,KEVADYFSTPEDESIDFLSTOMPYKII,PSGGI 121
QY 138 NNFPLYEKTAKKGRPWVISSGNOSQMDTMQVQIVKPLNPN---FCLQCTSAIYLOE 193
Db 132 TNLPLYEKTGRQAKVILSTGMVND---EIHQAVCI,LOENGTTDILSI,HCETEYDTPY 178
QY 134 DVNLVISHXOCLFPDIPIGYSGHETGIAISVAALV,GAKT,ERHITLIDKTMKGSHSAS 253
Db 179 ALINVLVTLKKEFPULTIGYSDHSGEVPLAALAMGL,ELIEKHTLIDNEMGBPHKAS 238
QY 254 LEPGEIABLVRSVRLVERALGSPYQOLLECEBA,CEKL--GKSVAKVKI,PSBTIITMDM 311
Db 239 ATPDILALVGVRIEVOIGLKEKE--DEBEVVRNKI,IAEKSIYAKKAI,AKGEVTEEN 296
QY 312 LTVXKGEF-KAYPREDIENLVGKRVLVYAEEDDTIMEELV,ND 352
Db 297 TTVK--RPGNGAISPMWKVYLVQVGSQDPEEDONICHSAPEN 336

```

RESULT 5  
ITS-70-15

```
068-09-03
Sequence 10910, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10910
LENGTH: 311
TYPE: PRF
ORGANISM: Streptomyces avermitilis
US-10-156-761-10910
```

	Query Match	Similarity	27.8%	Score 524	DB 14	Length 311
	Best Local	Similarity	36.9%	Pred. No. 2.9e-44		
	Matches	Conservative	53	Mismatches 115	Indels 24	Gaps 7
Qy	14	GGCHPFCIIAIGIQNHQGBLDVAKEMIRNAECGADCAATFOKSELEFKENREA--LERPY	71			
Db	15	GGCHPYYVVGELGINNGELENNAFKLIDAAEAAGCDAVFRTKEITCPROMOIER--	72			
Qy	72	TSGHMGK-TYGEHKHLEFSDHQVELQRYAEAVCIIFTASGMDEMAVEFLHEINVPF	130			
Db	73	--DTPMGRMTIYDIRHRVFEFGEDGYQIDETAYASKNIDIFASPMDTEAAVLEKPDIAPH	130			
Qy	131	KYGSGDNNPFYLEKTAACKGPWIVISSGMSQSMOTMGOVQIVPL-NPVFCFLQCTSAFP	189			
Db	131	KASASLTDELRLALRGTRTYILSTG--MSTPQIHANAVYLGSDDLCHANSYTP	187			

[illegible]

RESULT 6  
US-09-930-440B-B

```

1      ? Sequence 8, Application US/09930440B
2      ? Patent No. US20020142386a1
3      ?
4      ? GENERAL INFORMATION:
5      ?
6      ? APPLICANT: Betenbough et al.
7      ? TITLE OR INVENTION: Engineering Intracellular Signaling Pathways
8      ?
9      ? FILE REFERENCE: P2509P2
10     ?
11     ? CURRENT APPLICATION NUMBER: US/09/930,440B
12     ?
13     ? CURRENT FILING DATE: 2001-08-16
14     ?
15     ? PRIOR APPLICATION NUMBER: 60/227,579
16     ?
17     ? PRIOR FILING DATE: 2000-08-25
18     ?
19     ? PRIOR APPLICATION NUMBER: 09/516,793
20     ?
21     ? PRIOR FILING DATE: 2000-03-01
22     ?
23     ? PRIOR APPLICATION NUMBER: 60/169,624
24     ?
25     ? PRIOR FILING DATE: 1999-12-08
26     ?
27     ? PRIOR APPLICATION NUMBER: 60/122,582
28     ?
29     ? PRIOR FILING DATE: 1999-03-02
30     ?
31     ? NUMBER OF SEQ ID NOS: 8
32     ?
33     ? SOFTWARE: PatentIn Ver. 2.1
34     ?
35     ? SEQ ID NO: 8
36     ?
37     ? LENGTH: 346
38     ?
39     ? TYPE: PRT
40     ?
41     ? ORGANISM: Escherichia coli
42     ?
43     ? US-09-930-440B-8

```

Query Match	27.1%	Score 510.5	DB 9	Length 346
Best Local Similarity	36.3%	Pred. No. 8e-43		
Matches 123	61	Mismatches 140	Indels 15	Gaps 6

Qy 20 TIAIAGN0G0DULVAKRRIEMAE0AD0AKIY0NSLBERNF0K0M0AEIY0L0G0G0  
Db 5 YVAI0G0NNG0SVDIAREMILIKKAE0AVN0VKE0TF0ADKLIS1ARAE0I0N0T0E0L 64  
Qy 79 KTYGKH0LHESHQYREL0RYAE0VGIPTFAS0MD0MAVEFLH0LV0PFEKYVSG0DN 138  
Db 65 ES0L0MTK0LEKRYDYLH0MEYAN0SLND0PSP0FEDDSID0FLASL0K0IMK0IPSGELL 124  
Qy 139 NEPYLEKTAKK0RP--WV1SSG0WMDTK0V0Y0V--KLANPFCLO0TSK0Y0LP0 192  
Db 125 NBPYLEK0IAKPI0PKK0IIS0G0ATID0IK0S0V0I0FN0K0P0V0GN1IT1L0NT0Y0P0P 184  
Qy 199 EDVNL0RVISE0Y0LPP0DPI0GYSG0ERT0A10SVA0V0AL0AK0L0E0H1TLD0TKY0GSD0SA 252  
Db 185 EDVNL0M0AINDL0K0HP0K0N0IG0SD0MS0GF0Y0A10AV0P0G1TFIE0HFTL0DKMS0GD0HLA 244  
Qy 253 SE0P0EL0AELV0SV0L0VER0AL0SPTK0LL0PC0MA0CN0E0K0G0SV0AK0V0K1P0E0T1L0M0L 312  
Db 245 STE0P0ELK0L0IC0V0C0VER0SL0G0S0N0KV0Y0AS0ER0KN0K1I0V0AR0K0I1AK0T0ELIK0E0V0SE0KNI 304  
Qy 313 TYK0G0EP--K0VP0ED1P0NL0VG0K0VL0Y0E0ED0T10E0ELV 350  
Db 305 TTK--R0P0G0I0SP0M0E0N1I0L0G0----10E0Q1P0I0D0EL 336

## RESULT 7

US-10 106-698-6926  
; Sequence 6926, Application US/10106698  
; Publication No. US20030109690A1

```

GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA0005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 6926
LENGTH: 123
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (83)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (87)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (111)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (112)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (121)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6926

Query Match      20.2% Score 381; DB 14; Length 123;
Best Local Similarity 93.7%; Pred. No. 2,2e-30;
Matches 74; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Cy 183 OCTSAVPLQEDVNLVRISEYQKLPDIPIGYSGHETGIALSVAAVALGAKYLERHTITLD 242
Db 24 KCTSAIPLOPEDNLVRISEYQKLPDIPIGYSGHETGIALSVAAVALGAKYLERHTITLX 83
QY 243 KTWKGSDBSASLEPGEIAB 261
Db 84 KTWKGSDBSASLEPGEIGS 102

RESULT 8
US-09-816-028A-35
Sequence 35; Application US/09816028A
Patent No. US20020042369A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/09/816,028A
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:

```

```

OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-09-816-028A-35

Query Match      17.8% Score 336.5; DB 9; Length 346;
Best Local Similarity 29.1%; Pred. No. 3.7e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Cy 21 IIAIGONHOGDLVAKMIRMAKCGADCAKFEKSELEFKFNKALE-PPYTSKHSWGK 79
Db 19 VPELGINHNSLELAKIMVDAKSTAKIKKHQTHIVEDMSKAAKVIIPGNAKIS--- 75
QY 80 TYGHRKRLERFSDQRYEQLRYAEVGIPTLASGDMANAEFLHNLVPPFKVSGDTNN 139
Db 76 -IYEMCKCALDYDELBALKEYTEKLGIVLSTPSPAGANRLDMGVSAFKISGECNN 134
QY 140 FPYLEKTKAKGRPMVVISGQMSMDTKQVQIVKPLNPFCLQCTSAVPLQEDVNLV 199
Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILLDNIPFLMHTNLPPLHNLVRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIALSVAAVALGAKYLERHTITLDKTKWGSDBSASLEPGEI 259
Db 195 MLEIKKEF-SCWVGISDHTDNLACLGAVALGACVLRHFPDSWHRSGPDIVCSMDTOL 253
QY 260 AEL-VRSVRLVRLALGSPTRQLLPCENACNEKLGKSVAVAKYIPEGITLITMDMLTVKGE 318
Db 254 KELIIQSQAIMEGNNSKRAAKQOVTIDPARASVSTIDIKKGEVLMDNIMVWRPG 313
QY 319 PKAYPPEDIFNLVGRKVLVTVEDDTIMEE 348
Db 314 LGGISAEFENILGKKAALDIENDTOLSYE 343

RESULT 9
US-10-303-161-35
Sequence 35; Application US/10303161
Publication No. US20030148459A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,161
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-161-35

Query Match      17.8% Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 3.7e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Cy 21 IIAIGONHOGDLVAKMIRMAKCGADCAKFEKSELEFKFNKALE-PPYTSKHSWGK 79
Db 19 VPELGINHNSLELAKIMVDAKSTAKIKKHQTHIVEDMSKAAKVIIPGNAKIS--- 75
QY 80 TYGHRKRLERFSDQRYEQLRYAEVGIPTLASGDMANAEFLHNLVPPFKVSGDTNN 139
Db 76 -IYEMCKCALDYDELBALKEYTEKLGIVLSTPSPAGANRLDMGVSAFKISGECNN 134
QY 140 FPYLEKTKAKGRPMVVISGQMSMDTKQVQIVKPLNPFCLQCTSAVPLQEDVNLV 199
Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILLDNIPFLMHTNLPPLHNLVRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIALSVAAVALGAKYLERHTITLDKTKWGSDBSASLEPGEI 259
Db 195 MLEIKKEF-SCWVGISDHTDNLACLGAVALGACVLRHFPDSWHRSGPDIVCSMDTOL 253
QY 260 AEL-VRSVRLVRLALGSPTRQLLPCENACNEKLGKSVAVAKYIPEGITLITMDMLTVKGE 318
Db 254 KELIIQSQAIMEGNNSKRAAKQOVTIDPARASVSTIDIKKGEVLMDNIMVWRPG 313
QY 319 PKAYPPEDIFNLVGRKVLVTVEDDTIMEE 348
Db 314 LGGISAEFENILGKKAALDIENDTOLSYE 343

```

Db 76 -IYEWKCALDYKDELALKEYTEKGLVYLSTPFRAGANRLLEDWGSAPFKISGECNN 134  
 Qy 140 FPLEKTAKKGRPMWISSGQSMQMDTKQVQIVKPLNPNCFLOCSAYPLQPEDVNLAV 199  
 Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILDNBIPVLMHTTNLYPTPHNLVRLNA 194  
 Qy 200 ISEYOKLFPDIPIGSGHETGIAISVAANAVALGAKVLERHITLDTKWKSGDSHSLPGBEL 259  
 Db 195 MELKKEF-SCWVGJSDHTDNLACGAVNLGACVLERHPTDMSHRSGBPDIQCSMDTOAL 253  
 Qy 260 AEL-VRSVRLVERALSSPTQOLLPCEMACNEKIGKSVAAVKIPREGTILTMNLTIVYGE 318  
 Db 254 KELIISOEWAINRGNNESKKAQKQOVITIDFAPASVSIKDIKKEVLSMDNIWVRPG 313  
 Qy 319 PKAYPPEDIFNLGKRVLTVEEDDTIMEE 348  
 Db 314 LGGISAEPENILGKRLRDIENDTQLSYE 343

RESULT 10  
 US-10-303-118-35  
 ; Sequence 35, Application US/10303118  
 ; Publication No. US20030157655A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gilbert, Michel  
 ; APPLICANT: Makarchuk, Warren W.  
 ; APPLICANT: National Research Council of Canada  
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
 ; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
 ; FILE REFERENCE: 019633-000111US  
 ; CURRENT APPLICATION NUMBER: US/10/303,118  
 ; PRIOR FILING DATE: 2002-11-21  
 ; PRIOR APPLICATION NUMBER: US/09/816,028  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: US 60/118,213  
 ; PRIOR FILING DATE: 1999-02-01  
 ; PRIOR APPLICATION NUMBER: US 09/495,406  
 ; PRIOR FILING DATE: 2000-01-31  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 35  
 ; LENGTH: 346  
 ; TYPE: PR  
 ; ORGANISM: Campylobacter jejuni  
 ; FEATURE:  
 ; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 84  
 ; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
 US-10-303-118-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;  
 Best Local Similarity 29.1%; Pred. No. 3.7e-25;  
 Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNOGQDLDVAKKIMIRNAKCGADCAKFKQSELEFFKFKALF-RPYTSKSHWGK 79  
 Db 19 VPEIGINHGSLKAKIMVDAFSTGAKIKHQTHTIVEDMSYAAKXVIFGNKXIS--- 75  
 Qy 80 TYGHNKRLFSHOYRELQRYAEVGIFFTSAGDDEAVFELHBLNVPFKVSGDTNN 139  
 Db 76 -IYEWKCALDYKDELALKEYTEKGLVYLSTPFRAGANRLLEDWGSAPFKISGECNN 134  
 Qy 140 FPLEKTAKKGRPMWISSGQSMQMDTKQVQIVKPLNPNCFLOCSAYPLQPEDVNLAV 199  
 Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILDNBIPVLMHTTNLYPTPHNLVRLNA 194  
 Qy 200 ISEYOKLFPDIPIGSGHETGIAISVAANAVALGAKVLERHITLDTKWKSGDSHSLPGBEL 259  
 Db 195 MELKKEF-SCWVGJSDHTDNLACGAVNLGACVLERHPTDMSHRSGBPDIQCSMDTOAL 253  
 Qy 260 AEL-VRSVRLVERALSSPTQOLLPCEMACNEKIGKSVAAVKIPREGTILTMNLTIVYGE 318  
 Db 254 KELIISOEWAINRGNNESKKAQKQOVITIDFAPASVSIKDIKKEVLSMDNIWVRPG 313

Qy 319 PKAYPPEDIFNLGKRVLTVEEDDTIMEE 348  
 Db 314 LGGISAEPENILGKRLRDIENDTQLSYE 343

RESULT 11  
 US-10-303-128-35  
 ; Sequence 35, Application US/10303128  
 ; Publication No. US20030157656A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gilbert, Michel  
 ; APPLICANT: Makarchuk, Warren W.  
 ; APPLICANT: National Research Council of Canada  
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
 ; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
 ; FILE REFERENCE: 019633-000111US  
 ; CURRENT APPLICATION NUMBER: US/10/303,128  
 ; PRIOR FILING DATE: 2002-11-21  
 ; PRIOR APPLICATION NUMBER: US/09/816,028  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: US 60/118,213  
 ; PRIOR FILING DATE: 1999-02-01  
 ; PRIOR APPLICATION NUMBER: US 09/495,406  
 ; PRIOR FILING DATE: 2000-01-31  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 35  
 ; LENGTH: 346  
 ; TYPE: PR  
 ; ORGANISM: Campylobacter jejuni  
 ; FEATURE:  
 ; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 84  
 ; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
 US-10-303-128-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;  
 Best Local Similarity 29.1%; Pred. No. 3.7e-25;  
 Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNOGQDLDVAKKIMIRNAKCGADCAKFKQSELEFFKFKALF-RPYTSKSHWGK 79  
 Db 19 VPEIGINHGSLKAKIMVDAFSTGAKIKHQTHTIVEDMSYAAKXVIFGNKXIS--- 75  
 Qy 80 TYGHNKRLFSHOYRELQRYAEVGIFFTSAGDDEAVFELHBLNVPFKVSGDTNN 139  
 Db 76 -IYEWKCALDYKDELALKEYTEKGLVYLSTPFRAGANRLLEDWGSAPFKISGECNN 134  
 Qy 140 FPLEKTAKKGRPMWISSGQSMQMDTKQVQIVKPLNPNCFLOCSAYPLQPEDVNLAV 199  
 Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILDNBIPVLMHTTNLYPTPHNLVRLNA 194  
 Qy 200 ISEYOKLFPDIPIGSGHETGIAISVAANAVALGAKVLERHITLDTKWKSGDSHSLPGBEL 259  
 Db 195 MELKKEF-SCWVGJSDHTDNLACGAVNLGACVLERHPTDMSHRSGBPDIQCSMDTOAL 253  
 Qy 260 AEL-VRSVRLVERALSSPTQOLLPCEMACNEKIGKSVAAVKIPREGTILTMNLTIVYGE 318  
 Db 254 KELIISOEWAINRGNNESKKAQKQOVITIDFAPASVSIKDIKKEVLSMDNIWVRPG 313

Qy 319 PKAYPPEDIFNLGKRVLTVEEDDTIMEE 348  
 Db 314 LGGISAEPENILGKRLRDIENDTQLSYE 343

RESULT 12  
 US-10-303-134-35  
 ; Sequence 35, Application US/10303134  
 ; Publication No. US20030157657A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gilbert, Michel  
 ; APPLICANT: Makarchuk, Warren W.  
 ; APPLICANT: National Research Council of Canada  
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of

```

; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1998-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-134-35

Query Match          17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 3.7e-25; Indels 7; Gaps 4;
Matches 96; Conservative 63; Mismatches 164;

QY 21 ITAIGONHGDLDVAKRMIRAKECGADCAKPOKSELEFFKRNKALE-RPYTSKHSWGK 79
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 19 VVPEIGINNGSLDLAKIMVDAFSTGAKIHKQTHIVEDMSKAKKVIKGNKXIS--- 75
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 80 TYGKHKHLFESHDQYRELQRYAEVGIFFTASGMDMAVEFLHNLVPEFKVSGDTNN 139
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 76 -IYIMQKCALDYDELAKIYETKGLVYLSFSSAGARLEDMGVSAFKISGSCNN 134
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 140 FPLEKTKKGRPNVIVSGQSDTKQYQIVYVPLNPFCLQCTSAVPLQPEDVNLRV 199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 135 MELKKEF-SCWGLSDHTDNLACGAVAGACVLEHFTDSMHRSGPDIVCSMDTQAL 253
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAVALGAKVLESHITLDTWKSGDSASLEPGL 259
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 195 MELKKEF-SCWGLSDHTDNLACGAVAGACVLEHFTDSMHRSGPDIVCSMDTQAL 253
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 260 AEL-VRSVTLVERALGSEPTKQLPCBMACNEKLGKSVYAKKIPBGTITLMDLTVKGE 318
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 254 KELTIQSEGMALIKRGNESKKAQKQVITIDFAPASVVISIKIKKEVLSMDINWVRPG 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 319 PKAYPREDIFNLGKXVLTVVEEDDTIMEE 348
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
US-10-303-162-35
; Sequence 35, Application US/10303162
; Publication No. US20030157658A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Makarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1998-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346

```

```

; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-162-35

Query Match          17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 3.7e-25; Indels 7; Gaps 4;
Matches 96; Conservative 63; Mismatches 164;

QY 21 ITAIGONHGDLDVAKRMIRAKECGADCAKPOKSELEFFKRNKALE-RPYTSKHSWGK 79
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 19 VVPEIGINNGSLDLAKIMVDAFSTGAKIHKQTHIVEDMSKAKKVIKGNKXIS--- 75
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 80 TYGKHKHLFESHDQYRELQRYAEVGIFFTASGMDMAVEFLHNLVPEFKVSGDTNN 139
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 76 -IYIMQKCALDYDELAKIYETKGLVYLSFSSAGARLEDMGVSAFKISGSCNN 134
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 140 FPLEKTKKGRPNVIVSGQSDTKQYQIVYVPLNPFCLQCTSAVPLQPEDVNLRV 199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 135 MELKKEF-SCWGLSDHTDNLACGAVAGACVLEHFTDSMHRSGPDIVCSMDTQAL 253
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAVALGAKVLESHITLDTWKSGDSASLEPGL 259
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 195 MELKKEF-SCWGLSDHTDNLACGAVAGACVLEHFTDSMHRSGPDIVCSMDTQAL 253
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 260 AEL-VRSVTLVERALGSEPTKQLPCBMACNEKLGKSVYAKKIPBGTITLMDLTVKGE 318
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 254 KELTIQSEGMALIKRGNESKKAQKQVITIDFAPASVVISIKIKKEVLSMDINWVRPG 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 319 PKAYPREDIFNLGKXVLTVVEEDDTIMEE 348
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
US-09-882-227-490
; Sequence 490, Application US/09882227
; Publication No. US2003015836A1
; GENERAL INFORMATION:
; APPLICANT: Kleantchous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US2003015836A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 490
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-490

Query Match          16.1%; Score 103; DB 10; Length 340;
Best Local Similarity 27.7%; Pred. No. 9.1e-22; Indels 22; Gaps 10;
Matches 91; Conservative 66; Mismatches 145;

QY 18 PCFIATIGONHGDLDVAKRMIRAKECGADCAKPOKSELEFFKRNKALERTPT 72
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 4 PKIYAEELSANNGQDLNLAKESTLHAKESGADPVYLQTYTSPSCMTLNSXEDPFIIOGTLW 63
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 73 SHSWSKTYGKHRLFESHDQYRELQRYAEVGIFFTASGMDMAVEFLHNLVPEFKV 132
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 64 DRENLYELYQKASTPLEW-HAEFLRLAKL-DLGIIF--SSPSSQALSHLESINCPMKI 119
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



Thu May 13 11:53:08 2004

us-09-930-440b-6.rapb

Page 7

QY 133 GSGDTNNPYLEKTAKKGRPMVVISGMSQMDTKQYQIVKPINPF--CELOCTSAVPL 130  
DB 120 ASFEIVDLIEKAKRKQKPIILSSGIAITHELODAISLCRGVN-NFDITLTKCVSAVPS 178  
QY 191 QPBDVNLRISEYOKLPDIPPIGYSGHETGIAISVAALGAKVLERHITLDTKWSGDH 250  
DB 179 KIBDANLISVVKGEIF-GVYFEGSDHTIGSLCPILATTLGASMIERHFLINKSLQTPDS 237  
QY 251 SASIEPELAEIVRSVLAVERALGSPKOLLPCGMACNEKLGKSVAKVAPIPGTILT-- 308  
DB 238 AFSDFNGFKSMVAIKQSVLALGEEPRINPKTLERRRPFASLFFVTKIOQGEALTEN 237  
QY 309 -MDMLTVKVG-EPKAYPPEDIFNLVGRK 334  
DB 298 NIKALRPNLGLHPRFKYK-----ELIGQK 320

RESULT 15  
US-10-335-977-7626  
Sequence 7626, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7626:

SEQUENCE CHARACTERISTICS:

LENGTH: 340 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

HYPOTHEICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...340

SEQUENCE DESCRIPTION: SEQ ID NO: 7626:

US-10-335-977-7626

Query Match 15.2%; Score 287; DB 12; Length 340;  
Best Local Similarity 27.8%; Pred. No. 3,8e-20;  
Matches 92; Conservative 68; Mismatches 137; Indels 34; Gaps 14;

QY 21 IIAIGQNHODLVAKKIMAKBCGADCAKFK-----SELEFKFNRKALERPYSKH 75

DB 7 IVALISANENODINLAKESILHAIKESGADPFVKIQTVTPSCMTLDSK-----EDPFILOG 60  
QY 76 S-WGR--TYGHEKR--HLEPSHDQYRELQRYAEVGIFFTASGMDMAVEFLAEINVPF 129  
DB 61 TLMDKENLYGLYQKASTPLEW-HAEIPELAKL-DLGIF--SEPPSKALELLESIDCTM 116  
QY 130 FKVSGDTNNPYLEKTAKKGRPMVVISGMSQMDTKQYQIVKPINPF--CELOCTSA 187  
DB 117 YKIASFEIVDLIEKAKRKQKPIILSSGIAITAELODAISLCRGVN-NFDITLTKCVSA 175  
QY 188 YPLQPDVNLRISEYOKLPDIPPIGYSGHETGIAISVAALGAKVLERHITLDTKWSGDH 247  
DB 176 YPEKIDAEHLISVVKGEIF-GVYFEGSDHTIGSLCPILATTLGASMIERHFLINKSLQ 234  
QY 248 SDSASLEKGEIAELVRSVLAVERALGSPKOLLPCGMACNEKLGKSVAKVAPIPGTIL 307  
DB 235 PDSAFMDTFNGFQSMVGAIKQSVLALGEEPRINPKTLERRRPFASLFFVTKIOQGEALT 234  
QY 308 TMD--MLTVKVG-EPKAYPPEDIFNLVGRK 334  
DB 295 TSDNIRALRPNLGLHPRFKYK-----ELIGQK 320

Search completed: May 6, 2004, 09:25:07  
Job time : 40.8106 secs

Thu May 13 11:53:11 2004

us-09-930-440b-7.rge

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 4297.94 Seconds

(without alignments)  
10679.608 Million cell updates/sec

Title: US-09-930-440B-7

Perfect score: 1059

Sequence: 1 atgagtaataatataatcgt.....aatgagaacaaatattatg 1059

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3470272 segs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: gb\_vt.\*  
16: em\_da.\*  
17: em\_fun.\*  
18: em\_hum.\*  
19: em\_in.\*  
20: em\_om.\*  
21: em\_ov.\*  
22: em\_or.\*  
23: em\_pac.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hg\_hum.\*  
31: em\_hg\_inv.\*  
32: em\_hg\_other.\*  
33: em\_hg\_mus.\*  
34: em\_hg\_pln.\*  
35: em\_hg\_rtd.\*  
36: em\_hg\_mam.\*  
37: em\_hg\_vrt.\*  
38: em\_sy.\*  
39: em\_hg\_hum.\*  
40: em\_hg\_mus.\*  
41: em\_hg\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	100.0	1781	1	ECU05248
2	520	49.1	12105	1	AF361371 Escherich
3	328.6	31.1	17596	1	AF355776 Streptoco
4	328	30.8	16448	1	AF337958 Streptoco
5	326.4	30.7	17276	1	AF163833 Streptoco
6	324.8	30.7	10206	1	AB017355 Streptoco
7	324.8	30.7	18239	1	AF349539 Streptoco
8	324.8	30.7	21365	1	AB014245 Streptoco
9	324.8	30.7	25021	1	AB028896 Streptoco
10	324.8	30.7	95596	6	AX602206 Streptoco
11	324.8	30.7	128050	1	SA676849 Streptoco
12	321.4	30.3	1026	6	AX607023 Streptoco
13	312	29.5	9987	1	AB050723 Streptoco
14	300.6	28.4	26335	1	CJ11168X5 Streptoco
15	294.2	27.8	13390	1	AY102621 Streptoco
16	275.4	26.0	11938	1	AE011338 Streptoco
17	214.4	20.2	1849	1	AF195054 Streptoco
18	210.8	19.9	300171	1	AB015942 Streptoco
19	200.4	18.9	301235	1	AB016799 Streptoco
20	195	18.4	349562	1	BX569690 Streptoco
21	182.4	17.2	32661	1	LPY7311 Streptoco
22	170.6	16.1	12858	1	AB007719 Streptoco
23	166.2	15.7	14561	1	U67549 Streptoco
24	166.2	15.7	110000	6	AB271569_04 Streptoco
25	137.4	13.0	300254	1	AB016933 Streptoco
26	131.6	12.4	3876	1	AY034084 Streptoco
27	130.4	12.3	300000	1	AP005073 Streptoco
28	122.6	11.6	1388	1	CJA00855 Streptoco
29	118	11.1	3071	1	AF195055 Streptoco
30	118	11.1	4668	1	AF257460 Streptoco
31	116.4	11.0	13484	1	AF400047 Streptoco
32	116.4	11.0	13484	1	AY044156 Streptoco
33	116.4	11.0	282183	1	CJ11168X4 Streptoco
34	114.8	10.8	11474	1	AF215659 Streptoco
35	114	10.8	11442	1	AF400048 Streptoco
36	113.2	10.7	11455	1	AY044868 Streptoco
37	110	10.4	4749	1	AF305571 Streptoco
38	110	10.4	11474	1	AF167344 Streptoco
39	109	10.3	10292	1	AB011339 Streptoco
40	108.4	10.2	12390	1	AF401528 Streptoco
41	108	10.2	5067	1	AF126256 Streptoco
42	107.6	10.2	1050	1	U40740 Neisseria m
43	107.6	10.2	4176	1	X78068 N. meningitid
44	107.6	10.2	4908	1	NM5053 Neisseria m
45	107.6	10.2	15668	1	AB002366 Neisseria

#### ALIGNMENTS

RESULT 1  
ECU05248  
LOCUS ECU05248  
DEFINITION Escherichia coli polysialic acid gene cluster region 2 (neud and  
neub) genes, complete cds.  
ACCESSION U05248  
VERSION U05248.1 GI:454079  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
Aminozaki, P.W., Wright, L.F., Vann, W.F. and Silver, R.P.  
1 (bases 1 to 1781)  
Nucleotide sequence and genetic analysis of the neod and neub genes

Thu May 13 11:53:11 2004

us-09-930-440b-7.rge

Page 2

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL  
Immunology, University of Rochester, 601 Elmwood Ave, Rochester, NY 14642, USA

FEATURES  
source  
1.1781  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K1"  
/isolate="RS218"  
/db\_xref="taxon:562"  
/clone="PSR321"  
103..728  
/gene="neud"  
103..726  
/gene="neud"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAC43301.1"  
/db\_xref="GI:454080"  
/translation="MSKLTITGAGGSEKSIIDSLNHKVELIGIPIDKSKYQSYV  
ILGNDIADENKDYVYFIIIGKSTRHVNIIRKHLRLINIDKALISPNILS  
DEIFLGRKCIINLDRIDHVAIVIRNIEHNGECCSNISTVAVLGDVSELET  
VGSYTVVNGQLILGSKSIIGSGSVIRNIPSNVVAAGPTPLIRNGE"  
723..1763  
/gene="neud"  
723..1763  
/gene="neud"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAC43302.1"  
/db\_xref="GI:454081"  
/translation="MSNIYIVAEIGCNHNSVDIAREMILKAKAGVNAVKQEFKAD  
KLISAIAPKAEVOIKRGELISOLEMTKLEKDYDYLHMEYAVSLIDVSEPFDE  
DSIDPLASIKQIKMKIPSEELNLPYLEKIALPIPDKKIILSTGMATIDKQSVI  
FNNKVPVGNITILICNTEYPTFEDVNLNAINDKKIPKRNIGFSDHSGPTAATA  
AVYGIPTLEKFTLDKSMGSDHLASLEPDELKILCIQVRCVERSLGNSKSVVASE  
RKKIIVAKRSIIAKTEIKKEVSEKNIITTKPENGISPMEWYMLGKIALRQDFIPDE  
LIHSEFKNGE"

# ORIGIN

Query Match 100.0%; Score 1059; DB 1; Length 1781;  
Best Local Similarity 100.0%; Pred. No. 1,4e-158;  
Matches 1059; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTAATATAATATCGTCGTAATGCTGCAACCATATGAGTGTATAT 60  
DB 723 ATGAGTAATATAATATCGTCGTAATGCTGCAACCATATGAGTGTATAT 782  
QY 61 GCAAGGAATGATATTAAGCCAAAGAGCCGGTGTATATGAGTGTATAT 120  
DB 783 GCAAGGAATGATATTAAGCCAAAGAGCCGGTGTATATGAGTGTATAT 842  
QY 121 TTTAAGCTGATTAATTAATTCAGCTATGACCTAAGGAGATCAATTAATAAAC 180  
DB 843 TTTAAGCTGATTAATTAATTCAGCTATGACCTAAGGAGATCAATTAATAAAC 902  
QY 181 ACAAGGAATGATATTAAGCCAAAGAGCCGGTGTATATGAGTGTATAT 240  
DB 903 ACAAGGAATGATATTAAGCCAAAGAGCCGGTGTATATGAGTGTATAT 962  
QY 241 TATCTCCATCTATGAAATGAGCAAGTGAATTAATTAATTAATTAATTAAT 300  
DB 963 TATCTCCATCTATGAAATGAGCAAGTGAATTAATTAATTAATTAATTAAT 1022  
QY 301 GACGAAGACTCTATGATTTTGAATCTTTGAAAACAAAATATGAAAATCCCTTCA 360

DB 1023 GACGAAGACTCTATGATTTTGAATCTTTGAAAACAAAATATGAAAATCCCTTCA 1082  
QY 361 GGTGAGTATATGATTTTACCGTATCTTTGAAAATATGCCAAGTTCGATCCCTGATTAAG 420  
DB 1083 GGTGAGTATATGATTTTACCGTATCTTTGAAAATATGCCAAGTTCGATCCCTGATTAAG 1142  
QY 421 AAAATATCATATCAACAGAAAGGCTATGATGAGATTAATAACAGTGTGTTCTAT 480  
DB 1143 AAAATATCATATCAACAGAAAGGCTATGATGAGATTAATAACAGTGTGTTCTAT 1202  
QY 481 TTTATTAATTAATAAGTTCGGTGTGATATTTACAAATTTACATTTGCAATTCGAAATAT 540  
DB 1203 TTTATTAATTAATAAGTTCGGTGTGATATTTACAAATTTACATTTGCAATTCGAAATAT 1262  
QY 541 CCAAGCCCTTTGAGATGTAACCTTAATGCTATTAATGATTTGAAAACAACTTCCT 600  
DB 1263 CCAAGCCCTTTGAGATGTAACCTTAATGCTATTAATGATTTGAAAACAACTTCCT 1322  
QY 601 AAGATTAACATAGGCTTCCTGATCAATTCAGCGGGTTTATGCAAGCTATTCGGCGGTG 660  
DB 1323 AAGATTAACATAGGCTTCCTGATCAATTCAGCGGGTTTATGCAAGCTATTCGGCGGTG 1382  
QY 661 CTTATGAAATTAATCTTTATTTGAAAACATTTTACCTTTAGTAAATCTATGTCGAGCCCA 720  
DB 1383 CTTATGAAATTAATCTTTATTTGAAAACATTTTACCTTTAGTAAATCTATGTCGAGCCCA 1442  
QY 721 GATCATTTGGCTCAATAGAACCTGATGATCACTGAAACATCTTTGAGGGGTCAAGGTGT 780  
DB 1443 GATCATTTGGCTCAATAGAACCTGATGATCACTGAAACATCTTTGAGGGGTCAAGGTGT 1502  
QY 781 GTTGAAAATCTTTAGGTTCAATATGATTAAGTGTTCACGTTTGAAAAGGAATTAATA 840  
DB 1503 GTTGAAAATCTTTAGGTTCAATATGATTAAGTGTTCACGTTTGAAAAGGAATTAATA 1562  
QY 841 ATCTGAGCAAGAAAGTCTATTTATGCTAAACAGAGATTAATAAAGGAGGATTTTTC 900  
DB 1563 ATCTGAGCAAGAAAGTCTATTTATGCTAAACAGAGATTAATAAAGGAGGATTTTTC 1622  
QY 901 GAAAAAATATATACACAAAAGAGCTGTATATGATTCGATCGGATGAGTGTATAT 960  
DB 1623 GAAAAAATATATACACAAAAGAGCTGTATATGATTCGATCGGATGAGTGTATAT 1682  
QY 961 TTTATGGTAAATTTGAGAGCAAGCTTATTCAGATGAATTAATCATAGCGAA 1020  
DB 1683 TTTATGGTAAATTTGAGAGCAAGCTTATTCAGATGAATTAATCATAGCGAA 1742  
QY 1021 TTTCAAAATCAGGGGGAATTAATGAGCAAAAATTTATTTG 1059  
DB 1743 TTTCAAAATCAGGGGGAATTAATGAGCAAAAATTTATTTG 1781

RESULT 2  
AF361371 12105 bp DNA linear BCT 21-JUN-2001  
LOCUS  
DEFINITION  
Escherichia coli O104 antigen gene cluster, complete sequence.  
ACCESSION  
AF361371  
VERSION  
AF361371.1 GI:14517800  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli  
Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
AUTHORS  
Wang, L., Briggs, C.E., Rothmund, D., Fratamico, P., Luchansky, J.B.  
TITLE  
Sequence of the E. coli O104 antigen gene cluster and  
Identification of O104 specific genes  
JOURNAL  
Gene 270 (1-2), 231-236 (2001)  
MEDLINE  
21297197  
PUBMED  
11404020  
REFERENCES  
2 (bases 1 to 12105)  
Wang, L., Briggs, C.E., Rothmund, D., Fratamico, P., Luchansky, J.B.

and Reeves, P.R.  
 Direct Submission  
 JOURNAL Submitted (14-MAR-2001) Department of Microbiology, The University  
 of Sydney, Sydney, NSW 2006, Australia  
 Location/Qualifiers

## FEATURES

1..12105  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 /note="serogroup: O104"  
 141..737  
 /gene="galP"  
 141..737  
 /gene="galP"  
 /codon\_start=1  
 /transl\_table=1  
 /product="glucose-1-phosphate uridylyltransferase"  
 /protein\_id="AAK64366.1"  
 /db\_xref="gi:14517801"  
 /translation="MNVQCGEPDGLGSHILCAPAIGDNPVYVLPVVIDDAADPL  
 RYINLMTARFNETGSOVLAKMPGDISETSVIQTEPLDRGKSVIEPLEKPDQ  
 PQLTDSIDMAVGRVYLSADIPLELRTOPGAWGIQLTDAIAELAKQSVDAIMTGD  
 SYDCGKMGVMAQFAFKYGLRNLKEGAKRNGIERLSE"  
 1098..1718  
 /gene="wcdK"  
 1098..1718  
 /gene="wcdK"  
 /codon\_start=1  
 /transl\_table=1  
 /product="wcdK"  
 /protein\_id="AAK64367.1"  
 /db\_xref="gi:14517802"  
 /translation="MKKKLIIIGAGFAKAVIDLDHNEYALGFPVDSLKSGHGGP  
 IVGHSLSPISTPTVYFIAIGDDPDLALMSQIRDLKLTINVIDMKAIVSTRSSIG  
 TCYIYGRMAIYCDSELDQVYINTRALIEGHNISYCTNISTRVAILNGVYIGERSF  
 IGSTVNGQIKIGSSTIVSGSVYIRIPIRVVAGTPTRLIRER"  
 1723..2763  
 /gene="nab"  
 1723..2763  
 /gene="nab"  
 /note="NeuAc condensating enzyme"  
 /codon\_start=1  
 /transl\_table=1  
 /product="nab"  
 /protein\_id="AAK64368.1"  
 /db\_xref="gi:14517803"  
 /translation="MSKIYVABIGCHNGDFNLAKMVEEAKKAGVDAVKEPOTFRAD  
 CLISKYAPAEYQIKVTGNCQSOEMTRKLEPYBEIKLESYAKELGIDVPETPDF  
 ESIDFLSRNOKMKRIPSGELINLPYEKISRLPEKRIIVLSTGMATIEETKAYVA  
 LTKGMAPEKTIICHTVEPTPEYEDNLSIISIKRFGKQVGGPSDMSHRSYATA  
 SVPGITRIRKHTLANKFRSGDHKASVTPDELALCEGIRAYEKLCSAKIYVTSR  
 OKRIYARKSIYAKCNIRKSEITITDITTRFNGISIPMTITDILGKAEKRFDEQ  
 LHSRSGFDEQEI"  
 2763..4034  
 /gene="nabC"  
 2763..4034  
 /gene="nabC"  
 /note="CMP-NeuAc synthetase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="nabC"  
 /protein\_id="AAK64369.1"  
 /db\_xref="gi:14517804"  
 /translation="MTNRIAIIPRAGSKGLPKRNVIMLIDKPLIATVIEAESQCF  
 NKIIVTSDLEYSIAERYAGEVITIESELSADPATYWTENTVLSHNDCCYPMLO  
 PTPSRNARHITKEAITLPBGVPAEPIVSMETSEISQIMKTLDSMRISNRNNDYSG  
 YRRODNQKYSFPCAIFITGYBEILKKKHFGADSLATYAKKADSIDIDRIDPEPAI  
 SRRLQNNKIKDEVAARIRKQFSITHAIKPTLLGHSTFDNWEISDLAGKSVNIG  
 VSGITSDYVNLIDNARIKEIGDVAFLGNDIYVIOGDENDTLRMWKIIDRIVK  
 LAPNVIIYFLSLPVGNGRVDRNNKVIIEHLNAFLPKGLDKINHVRVSLSDMYDDEGN  
 LSADPTVGLHFNKAVROLKEIMELIK"  
 4031..5194  
 /gene="nabA"  
 4031..5194

## CDS

4031..5194  
 /gene="nabA"  
 /note="GlucNAc-2-epimerase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="nabA"  
 /protein\_id="AAK64370.1"  
 /db\_xref="gi:14517805"  
 /translation="MKKILYVTSRAEYIMKRLLOOLKEDSDISLIVATGMECDK  
 YGMTKYIENDGPRVDKIDIEINNDNNAIITKMSCLSEFGIYFQDNKPDAMVILG  
 DRYEIMSYALAAAHNLPITPHGSEKTLGNYDEFTGHSITKMSLALASTHEPKRI  
 IOMGHPNSVYDIGALGANSLLTLPLDVSORFGLSKVYFVVFHFHTLGSVP  
 EIQVSLKSLSHETIKTNFVFGSDSDTQKTKDKPCENKNTKVFVSLIIEBY  
 LALCKYSQGLIGNSSGGLIEVPSLGVPTVINGRQKRGVAGATVINCSTEGISEAI  
 KKCLEDHYINGIKNVNPVYKHNANNAKIKKEIFLLSNHGYDFYLSFEL"  
 5279..6229  
 /gene="wbaA"  
 5279..6229  
 /gene="wbaA"  
 /note="putative glycosyl transferase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="wbaA"  
 /protein\_id="AAK64371.1"  
 /db\_xref="gi:14517806"  
 /translation="MKMNNVVPDPSPYCIITCYILRHOADPTIYSDNVSLRDIN  
 PPKERYIIRKGNKTSKITYFPCPLRLIINPKLISNRDQYRELGDHLEFSK  
 FLNEFTLEDELAVRYPHYSRLKYLIGHTRGRSRVKNKILISGMIDITPTINK  
 VEFEDLIAWELKMSVOQKEINHIENYICDELVAADWILALQPSDEGPISENKIRL  
 YDBIIEYDVKIYIRQHPRELTDYSYFKDVKNKVAPVELVVLVNSGPSIAVTLF  
 SGGIFNICEKRIFKGTSFSLSKHGNIINKEEVAKK"  
 6226..7338  
 /gene="wzy"  
 6226..7338  
 /gene="wzy"  
 /note="O antigen polymerase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="wzy"  
 /protein\_id="AAK64372.1"  
 /db\_xref="gi:14517807"  
 /translation="MTSYFTNRYVLAIFAAVHERINLFTSGVKEISIGVFCST  
 FLSHLKRDVGPDEYTYVYIKNVIAKRIIDITTYRGLSTLPSLDDGYVYRVY  
 VSPULPEFVRYILBKIPMGIIIVCFVSGYLDLRIRQVLAVALFVLSITKX  
 BLRFFAIITIASFFHLSALMLPVPYVNERFEPFLIIGRLVLYISGFTTBQ  
 ALIYNNIYVNEIYGNISQFALOGFGTGFVPHILIFVSIYNEKONYLNRLEFT  
 AGLLMIITAPGNLIMRFGQFYVSLIILALPLAISKIKREYSYNLLKCFLLFTFF  
 OSOGRNYSYOSIFSDFNSLSTPSSDRITINH"  
 7344..8567  
 /gene="wbaB"  
 7344..8567  
 /gene="wbaB"  
 /note="putative [glycosyl] transferase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="wbaB"  
 /protein\_id="AAK64373.1"  
 /db\_xref="gi:14517808"  
 /translation="MKNIILFTNNALSTGCAVYILNAKSLRNRGHVALLDRVDNI  
 HYKIDIGELIYVRDKIRIKVNSPTLNKNSQAFPTQNSVFPNFKILYAKKISFP  
 YELAPCEMAFFKEINTKKEKDAIINNNAVVDRIYBSKLSNNYVNFHNSIEIF  
 SRREFTLLPSKIFRNVKLSVSKJADBLILKGNDRKIIITIVNPFETLEGNS  
 RIOCEGNIPLKKEIIVTVSTLDRKVRDRITKAPKIRREGEIDLIIEGQLENDL  
 ONLKEKNIETKVEYFLQPNPYFIKAOALLISSDSGLPYIILISLIGTPLYST  
 DCPGPHILVNMWDSLVSLTSDSDNNI CNELAEKAILLCKKDKKVKKSDQ  
 RPKDKTISWEILF"  
 8549..9844  
 /gene="wzx"  
 8549..9844  
 /gene="wzx"  
 /note="O antigen flippase"  
 /codon\_start=1  
 /transl\_table=1

[illegible]

Db	2203	CTTCTTAAAAATGGAAATGGCTCTCGAAAAAAGATCACATATATTACATTGTAACTGATAT	2262
Qy	541	CCAAAGCCCTTTGAGAGATGTAAACCTTAATGCTATTATGATTGAAAAACACTTCCT	600
Db	2263	CTTCACACATATAGAAAGATGTAAATTTTAACTCAATTATGGCTTAAAGAAATGATTTAAA	2322
Qy	601	AAGATTAACATAGAGCTCTCTGATCATCTTACGCGGGTTTATGACGCTATTTGCGCGGTG	660
Db	2223	GGATATAAGCTGGCTTTTGTGACATTCACACAGGTATTATGACGCTATTGATCAGTT	2382
Qy	661	CCTTAGGAAATTAACCTTTTATGTAAAAACATTTCACTTTAGATTAATCTATGTCTGCGCCA	720
Db	2383	CCGATATGGAATTACATTCATCGAAAAACATTTTACTTAAATTAATAATTTTGAAGGGCCT	2442
Qy	721	GATCATTTGGCTCAATAGAAACCGATGAACTGAAACATCTTTGATTTGGGCTCAGGTGT	780
Db	2443	GATATAAAGGATGATTACACACAGATGACTACGATTATTGTGTAGCGGATTCGGGCT	2502
Qy	781	GTTGAAAAATCTTTAGGCTTCAAAATGTAAAGTGTTACAGCTTCAGAAAGAGATTA	840
Db	2503	GTTAAAAAAGTTTAGGAAGCTATGCGAAAAATAGTAACTGAACTGAAAAAATAAG	2562
Qy	841	ATCGTAGCAAAAGTCTATTATGTGTAAAAACAGATTAATAAAAGGTGAGCTTTTCA	900
Db	2563	ATATGTACCAAGAAAGTCCATATGTGGCAAAATGCAACTTAAAAAAGGTGAATATATACG	2622
Qy	901	GAATAAATATTAACAACAAAAAGACCTGTATGTATCAGTCCGATGAGAGTGTATPAT	960
Db	2623	ATTGATTAATATACCAACAAAGCGGCCACAGAAACGGATTGTCTATGTATGTATGAT	2682
Qy	961	TTATTTGGTAAATTTTGAGAGCAAGACTTTATTCAGATGATTTATATTTCATAGGAA	1020
Db	2683	ATCTTAGCAAGAAAGCGGAAAAAATTTGATATGAGATCAATTAATATAGTCATCTCGA	2742
Qy	1021	TTCAAAATATCGGGGGAATATGAGACAAATAATTA	1056
Db	2743	TTGATGAAACGAGAGATCTAATTATTAATGACTAATA	2778

FEATURES	source	REFERENCE	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	AUTHORS	TITLE	JOURNAL
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-10_signal	17596 bp	DNA	linear	BCT 29-APR-2001						
gene	-35_signal	17596 bp	DNA	linear	BCT 29-APR-2001						

```

CDS
    /gene="cpsA"
    /codon_start=1
    /note="similar to LtyR"
    /transl_table=11
    /product="CpsA"
    /protein_id="AAK43602.1"
    /db_xref="GI:13876770"
    /translation="MSNHSRROCKKSHSTPLRVNLFVLPILLSVSLFLMYRHF
    LAFRLNVTGVVTVLILASLFLCKIKARLEITLILVASIPVATLIEKSTIDL
    TNNLKTASYSIEIEMSVVPKRSKTNITANVSKLAPKPCDPSNTDILNHSKGI
    SITPKTDSYQAVNRIKNGDSQAVLNVAVSLBELSTPDRSQIKITITTKIKKI
    NKTNTNKGAVNIVYISGIDITGSGISTVSRDVNIIMVNTNTHKVLTLTPRDVVK
    IPDGGNOYDKLTAGLGVETSMKLENIYDINDYARINFSFLKIDLLGGVTV
    YNDQAFTSKHGNDPFGVQVTLNSEQALGFVEREYSIQGDNDRGMRKVIATINK
    LASSQVTKLNTSITQSLOTSVOTNMTINDININDINNSTGQRFVESQALTGHSGTG
    ELPSYMRPQALNMSIDSSLSNASKIKNTME"
    2388..3133
gene
    /gene="cpsB"
    /codon_start=1
    /note="cpsB"
    /transl_table=11
    /product="CpsB"
    /protein_id="AAK43603.1"
    /db_xref="GI:13876771"
    /translation="MIDISHIVFDVDDGPKTLESLISLIESYQGVNIIVSTHRR
    KMEETPDIIIFKNSIVHBAKREKREHQLDGGELYYTSDMLSKLKQKPLANT
    KRLIFPSQTSKQDITHTLSALNMTNGLTPVVAHIERVNALEOKREVEIIMGCYT
    QINSSHILKQKFNCHKRKRKRAVFLLENLVHFAVSDMHLNIVPPLLAAYKILC
    RDEKERNQLFENASILKNHYI"
    3132..3834
gene
    /gene="cpsC"
    /codon_start=1
    /note="cpsC"
    /transl_table=11
    /product="putative capsular polysaccharide chain length
    regulator CpsC"
    /protein_id="AAK43604.1"
    /db_xref="GI:13876772"
    /translation="NNKLAITEVEINIPNLKKMKKKELITVAIAFPAGLFSLP
    IVPQYTSSTRIVNPNPNNSITAGDQAGSFLANDYKEIITSTDVLEKVISSEKL
    NYPSSQLQKITISILKDRVISISVEDANPKWSQPLANSVREAAVSKIKAVAYEDI
    TLEKGNLPKAPSSPNIKNVLLIGFVAGLSTIVIVINGIIDDRTVEDIEKVLGT
    TSLGIVDLNKL"
    3833..4543
gene
    /gene="cpsD"
    /codon_start=1
    /note="cpsD"
    /transl_table=11
    /product="capsular polysaccharide chain length
    regulator/exporter CpsD"
    /protein_id="AAK43605.1"
    /db_xref="GI:13876773"
    /translation="MTRLAIVDSKLRQAKKTEEPNARITNIQSGKENKIATISVR
    EGGKSTTISIALSAGFTLLIDADRNVSMGTFATGTIKGLNYSAGNDL
    GDIICETNVPRLNVPSSKVPNPTALLQNAVFNKNIIBAKNIPYIILIDPPIGLV
    DAILIANACDGLITVQAGRIKKNYVEKMEQSGSKRLGILLKNVESVATYDY
    GGYNTGKDRKAK"
    4535..5944
gene
    /gene="cpsE"
    /codon_start=1
    /note="cpsE"
    /transl_table=11
    /product="putative glycosyltransferase CpsE"
    /protein_id="AAK43610.1"
    /db_xref="GI:13876778"
    4535..5944
CDS
```

```

RBS
    /gene="cpsF"
    /codon_start=1
    /transl_table=11
    /product="glucose-1-phosphate transferase CpsF"
    /protein_id="AAK43606.1"
    /db_xref="GI:13876774"
    /translation="WIVQVVVYFASLITLPLITPNKSKDLPVLLIHYIPLVDSF
    LRVDSRGVLEFKNVLEKISFTYIFISSLFFIRKNSFTTRLSPFETINSLIYV
    LNSFLKRYKRYSAKSRDITVILTNKDSLSKMTFRKTYDHTYIAVCLIDSEKDCY
    DLKNSLRIINKDLSELCTCLVDAQFINIPIELFGKYQIODIINDIEANGVIVAN
    VEAISFDNIGKRIQTFEGSVIYSMKPYVSHIARPLDITGAILGLICGVAI
    FLVPIRKDGGPAIISQNRVGNRIIFPYFRSRRVABOIKDOLLVNMGTGMLFK
    LEDPDRIRKIGKFLIKTSIDELPPVYVLKQDMSIVGRPPVDEYEKNSYOKRLS
    FKPTGTGMOJSGRNITIDPBEIYVLDVQVINEWSIMSDIKILLTLKAVILGTAK"
    5958..6417
gene
    /gene="cpsG"
    /codon_start=1
    /note="cpsG"
    /transl_table=11
    /product="beta-1,4-galactosyltransferase enhancer CpsG"
    /protein_id="AAK43607.1"
    /db_xref="GI:13876775"
    /translation="MKICLVSSGGLAHNLKPILEKEDRFVYTFDKEDARSILRE
    EIVHCFPTNRYNVLKTLIAKVLKREKRPDIYISSGAVAVPFVIGLFGCKT
    VYIYFDDIDKPTLGLKLYVPTDKEIVQWEMKKVYKALNIGIP"
    6402..6908
gene
    /gene="cpsIVg"
    /codon_start=1
    /note="cpsIVg"
    /transl_table=11
    /product="beta-1,4-galactosyltransferase CpsIVg"
    /protein_id="AAK43608.1"
    /db_xref="GI:13876776"
    /translation="MIFVTVGHEQGFRLIKREVRLKGTDAIDQEVFIQSGYDFEP
    ONCWSKRLSYDMSNVMKKEAIVITHGGPATFNMAVSKGKKEIVYVPROEGEYVNN
    HOYDVYNAKTMNFIYVDIERLQNVYEGTNMRPFLPTRNSNIEEFKYLKELCD
    ENQ"
    6882..7992
gene
    /gene="cpsIVH"
    /codon_start=1
    /note="cpsIVH"
    /transl_table=11
    /product="putative polysaccharide repeating unit
    polymerase CpsIVH"
    /protein_id="AAK43609.1"
    /db_xref="GI:13876777"
    /translation="WKINKNSFYAIFLVNPKRSIGEGNSTYKIYVAFILGCI
    KFLDLSYFERKVLITFLPLATILNLPFHAKVFTITLILFLLKQISLKKARSII
    IGSRLIGVLANQIFVKDLIETKYNFRDQPLIRSDLGFRHPPIHFRPVYVPLV
    VTLFPRKRLITIAIFLTLNLFVLYOYTSRSGYVYVLLFLIIVYKNNLIRKIFIV
    APYIOLPLAFLPFCSTIEFNSNFQKDSLITGRILANVQLVGLTLFGNSPKETS
    VNFDSYSLNMGVNLTPCMILIIYIYSKKVNVVEQLILFIMSVILFTSPSPSI
    VNISIMVPGKIFCGGVDDLQREFWTANKN"
    7875..7880
RBS
    /gene="cpsIVM"
    /codon_start=1
    /note="cpsIVM"
    /transl_table=11
    /product="putative glycosyltransferase CpsIVM"
    /protein_id="AAK43610.1"
    /db_xref="GI:13876778"
    7949..8800
gene
    /gene="cpsIVN"
    /codon_start=1
    /note="cpsIVN"
    /transl_table=11
    /product="putative glycosyltransferase CpsIVN"
    /protein_id="AAK43610.1"
    /db_xref="GI:13876778"
    7949..8800
CDS
```

QY	1	ATGAGGAAATTTATATATCGTGTGAAAATGTGTGCAACAAATATGAGAGTGTGATAT	60
Ds	13047	AAGATATATGCTTATATATATTGACAGAGATGTGGTGCATATCATATGAGATATATATCTT	1310
QY	61	GCAGAGAAATATATTTTAAAAGCCAAAGAGCCGGTGTATATGACAGTAAAATCCAAACA	120
Ds	13107	GGCAAAAAATGTGTGATGTGTCCTGTCTGTGTGTGTGTATGTCTGTATTAATTTGAGACT	1316
QY	121	TTTAAAGCTGATAAATTAATTTATTCGCTATATGCACTTAAGCGAGATATCAATTAATAAAC	180
Ds	13167	TTTAAAGCTGAAAACTTATTTCTTAAATTTGTCTCCAAAGCTGAAATATCAAAAAGCACT	1322
QY	181	ACAGAGAAATTTGAATCTCAGTTAGAAATGCAAAAAGAGCTTGAATTAAGATATGAGAT	240
Ds	13227	ACAGAGACGACGACAGGTCACTTATGAGATGACGAAAGCTTATAGAGTTAAGCTTTGAAGA	1328
QY	241	TATCTGCATCTATGGAATATGCACTCAGTTAAATTAATATGTTTCTTACCCCTTT	300
Ds	13287	TACTTGAAGATCGTGAATATATGCAATTTCAAAGGTGTGAGACCTTTTCACACCTTTT	1333
QY	301	GACGAGACTCTATATGATTTTATTGACATCTTGAAACAAAATAATGAAATAATCCCTTCA	360
Ds	13347	GATGAGAGCTCATATGATGTCTTATTTCTTCAAGACATGCCAATTTGACAAAATTCATCA	1344
QY	361	GGGAGATTATTAATTTTACCGATCTTGAAAAAATGACGAAGCTCCGATCCCTGATAG	420
Ds	13407	GGGAAATCACTAATTTACTTACTTATGAAAAGATGTGGCAAGCAAC-----AAAG	1344
QY	421	AAATATATCATATCAACAGAAATGCTCATATGTATGATTAACAAACATCTGTTCTAT	480
Ds	13458	AAAGTATTTCTTTCGACGGGTATGCGGTAAATGAGAAAGATCCATCAAGCGGGAATATT	1353
QY	481	TTATATTAATTAATAAGTTCCGGTGTGTATATTAACAATATACATTCGATACTGAAAT	540
Ds	13518	TTTGTGTCAAGATGTGCAACCG-----ATATTTCTAATTTTACATGTGTAACAAGATGAC	1353
QY	541	CCAGCCGCTTGAGAGATTAACCTTAATGATATTAATGATTGAAAAAAGCTTCCCT	600
Ds	13572	CCAAACACTTACCCCTCTCTAAATTTAAACGTTATCTACTTTGAAAAGATGAATTTAAA	1363
QY	601	AAAGATTAACATATAGCTTCTGTATCATTTCAACGGGTTTTATGCAAGTATGTGGCGCGGTG	660
Ds	13632	GATTTAACGATATGATTAATGCGATCATTCATATGAGATCAGAAAGTACCTAATTTGACAGACGA	1363
QY	661	CTTATGAGATAATCTTTATTTGAAAAAACAATTCACTTATAGATAAATCTATATGTCTGGGCCA	720

```

/note="similar to lytR"
/codon_start=1
/translate_table=1
/product="CpsA"
/protein_id="PAK11658.1"
/db_xref="gi:13022159"
/translation="MSNRSRKHQKSHHTPERVNLELVFPIILSVSLFIMYRHH
LAFSLNVIYGVVILLIILSLAFCIKRNKAFITLIILVSLISIVATLLYKFKSID

```

```

TNNLNKTAASYEISEMSVVPKDSKI TNIEAVSKLAPVKNQDTSNTIDLIEHKISEKGI
SITPQRTDSYDAVYRIKNGDSQAMVYLAIELEPPDEKSCIKTIYTYRIKKKI
NKKNTHKXGVNVIYSGIDTGSISTVSRSDVIMVNTNTHKLTITTPRDYAVK
IPDGNNOYDKLTHAGLVGEMTKLELYDINDVYARINPSSFLKIDLGKVTY
VNDQATSKXGNFDPFGVGTANSBEALGFVBERYSOGDNDRGNOBKVAALINK
LMSQSVTKNSITTSQLOTSVOTNMTINDINDLNQSTGQRTVESQALRGSGTG
ELPSYAMPQAOYMSIDSSLSNASKIKNTER"
gene      2415..3162
RBS       /gene="cpsb"
          2415..2421
          /gene="cpsb"
          2431..3162
          /gene="cpsb"
          /codon_start=1
          /transl_table=11
          /product="tyrosine phosphatase Cpsb"
          /protein_id="AAK1659.1"
          /db_xref="GI:13022160"
          /translation="MIDISHIVPDVDDGPKLTLESLIIESSVQGVRIIVSTSHR
          KGMFETPBDITFKNSITVYHEAKRFESHQILYGGELTYTSDMLEKIKQIPTANT
          KFLIFSWQTSWDITLALSNVLMGITPVVAHIERVNALENQKEREKELINMGCTY
          QINSHILKQKLFNDHQKFKKARARFLEENLVHFAVADMENLDVAPPLAAYKIC
          RDGKERANOLFENQAOSILKNHYI"
          3161..3863
          /gene="cpsc"
          3161..3167
          /gene="cpsc"
          3171..3863
          /gene="cpsc"
          /codon_start=1
          /transl_table=11
          /product="putative chain length regulator Cpsc"
          /protein_id="AAK1660.1"
          /db_xref="GI:13022161"
          /translation="MNRKIANTEVEINIFMLKKLWKKFLITFVAIAPFAGLPSLF
          IVPQYTSSTRIVINPTNPNNSITQADIQAGFLANDYKEIITSTVLEKVISSEKL
          NYSQLOKQITVSIILKTRVISISVEDANPKMSQKLANSVBAASKAVATQVED
          TLEKGNLPKAPSSPMKKXVNLGFIWAGLSITVIVINGILDRVNTBEDIEKVLGL
          TSGIVPDLNKL"
          3863..4563
          /gene="cpsd"
          3863..3868
          /gene="cpsd"
          3874..4563
          /gene="cpsd"
          /codon_start=1
          /transl_table=11
          /product="capsular polysaccharide chain length
          regulator/exporter Cpsd"
          /protein_id="AAK1661.1"
          /db_xref="GI:13022162"
          /translation="MTRLRIVDSKLRQAKKTEBYFPAIRTNIOFSGENKILATSVR
          EBEKSTSTSLALSIAQGFKLIDACFRSVMGSTRKATGILKGLINVSQNDL
          GDICETNPRILMVPSSCKPRPPTLLIONAYMKRILPAIKNIDYLIIDTPRIGIVY
          DAIIASACDGFVLVTOAGRIKRYEKAKQNGSGSKFLGIILAKVMSVATIDYD
          GNYGKDRRK"
          4577..5964
          /gene="cpsb"
          4577..4582
          /gene="cpsb"
          4615..5964
          /gene="cpsb"
          /codon_start=1
          /transl_table=11
          /product="glucose-1-phosphate transferase Cpsr"
          /protein_id="AAK1662.1"
          /db_xref="GI:13022163"
          /translation="MIGVYVYVSVSLTLLITPNKSNKDLFVLLIHVIVFYLDF
          YRDFSRGIEEFKMYLKISFYITFISSSIFPIPKASPTTTLSPFTPIANMSILLYL
          LNSFLYRKYSYAKSRDQKVVLLTNKSSLSMTNRKIDNHYIACVLDSSEKQY
          DLHNSLRINKDALTSBLTCLVDAFINIPILFGRYQIDIIIDISAMGVAVNV
          VEAISPDNIGEKRIQTFEGSVITYEMKRYKSLAKRFLDIMAGIIGLLCGIVAI
          FLVPOIRKDGPAIFQNRVGNRGRIFRFYKRSKVDQAIQIKDLVHNMQNTGLMFK

```

```

LADDERITKICKFIRKTSIDELPOFTYNVLKQDMSLVGRPPTVDEYKNTQKRLS
FKPGYTGMOQSGRRNITDPDEIVKLDVQYINEMSIWBDIKIILLKRVLLGTAK"
gene      5975..6437
RBS       /gene="cpsf"
          5975..5981
          /gene="cpsf"
          5988..6437
          /gene="cpsf"
          /codon_start=1
          /transl_table=11
          /product="beta-1,4-galactosyltransferase enhancer Cpsf"
          /protein_id="AAK1663.1"
          /db_xref="GI:13022164"
          /translation="MKITLVSSSGCHLAHLNLKPKMEKDRFWYTPKEDARSLIRE
          RIVHCFPPTRNKNLVKNTILAKVKLKERPDVITSSGAANVPFPIYIGLFGCKT
          VIVEVFRIDKFTLGLKLVYFVTDKFLVQWEEKKIYKRALNLGSI"
          6422..6910
          /gene="cpsg"
          6422..6428
          /gene="cpsg"
          6437..6910
          /gene="cpsg"
          /codon_start=1
          /transl_table=11
          /product="beta-1,4-galactosyltransferase Cpsg"
          /protein_id="AAK1664.1"
          /db_xref="GI:13022165"
          /translation="MIFPTVGTGHEQPNRLIKENDRLKGTGAIDQEVPIQNGSPFEP
          QNCQSKRLSTDQNSIMKEASIVITTHGSPATMSVIGLGLPVVPRPKQSGEIND
          HOIQPLNSIAHLVPLAMIEDVDGLAEALKRNIAVEKIQGNDMPCHLKKIIGEL"
          6898..8064
          /gene="cpsH"
          6898..6903
          /gene="cpsc"
          6907..8064
          /gene="cpsH"
          /codon_start=1
          /transl_table=11
          /product="capsular polysaccharide repeating-unit
          polymerase CpsH"
          /protein_id="AAK1665.1"
          /db_xref="GI:13022166"
          /translation="WRKTLDDYSLPYALMVLIVPNQYQFLITIVILLKXSEF
          RISINSSILFLMLFYLFAILIRGVEDITPQFIAELKLTISTGALPFTYRK
          ADPNSSVVRNKKVNVFVLFITVYLFEPNEFTFLCDLSEIMSPMSKRLATVP
          EYATLLQGFIFETVYILFLKOQRYGEMIFITLFCALVGARFLLOMIIILGLY
          LEIINKFNKLTCKAVPLIIGIILLVCFSYKVSILINYIHYRFQSSSTPLTVY
          ESIRIILGNFLIOGIRVPSSVGIPLGSHSYSISIPRTSPTGLFESSILLFLYRE
          AIKQRIITYKFLFGLILLMYFBERDPNHSVILLFTLGIYGRNDKKS"
          8034..9013
          /gene="cpsVII"
          8034..8040
          /gene="cpsVII"
          8045..9013
          /gene="cpsVII"
          /codon_start=1
          /transl_table=11
          /product="beta-1,3-glucosyltransferase CpsVII"
          /protein_id="AAK1666.1"
          /db_xref="GI:13022167"
          /translation="MIKKLVSVIVPVNSSEIVENCVESLLQOYPEIEIILLIDGST
          DKSSHICNPLKRDGRVYVAKNGASASANNVGLIEMAGEITIVDSGDVYALMIE
          IMENLILNEMADISIDPEVDDPYRKRRKRGYVYRONMSIKKEFPMSKENVWG
          KLYKSLIIGDLNFKNKYIGEDLLFNQILNKEHRIYVDTRSLSLYTIEKSTANQ
          FNKNTLIDIDIFNBIHQDSPELEFVYAKVREKIKLRNFTSLGELADENLRORI
          KFMQDIKSYISCKAIRFLSKKHICTLYLMKTFPVYIKYNNFKQK"
          9031..9752
          /gene="cpsVIJ"
          9031..9037
          /gene="cpsVIJ"
          9060..9752
          /gene="cpsVIJ"
          /codon_start=1

```



Query Match	31.0%;	Score 320;	DB 1;	Length 16448;
Best Local Similarity	58.4%;	Pred. No. 5.7e-43;		
Matches 618;	Conservative	0;	Mismatches 425;	Indels 15;
				Gaps 2

Db	13129	ATGTTCTCGTGAATCAGTAGTGGTCTTTAAACCAATTAAAAAGCGATATTATTCA	1318
0y	901	GAATAAAATATTAACACAAATAAGACCTGGTAATGTATCAGTCCGATGAGTGTATAT	960
Db	13189	ATAGAAATATTTACGGTAGAGGCGCCAGGTAAATGTATTCTCTTGAACATGGTATGAT	1324
0y	961	TTATTGGTAAATTTGACAGACAGACTTTATTCAGATGAATTAATTAATTCATAGGAA	1020
Db	13249	ATCTTGGACACAGAGGCGCAGATGATTTGAGAGAGATGAAGTATTCGATTCAGCG	1330
0y	1021	TTCAAAATCAGGGGCGATATATGACAAATAATTAT	1059
Db	13309	TTTGAAATTCATATGCCCTAGTATATTAACACTTAT	13346
RESULT 5			
AF163833		17276 bp	DNA
AF163833			Linear
LOCUS			BCT 03-AUG-20000
DEFINITION			Streptococcus agalactiae capsular polysaccharide synthesis locus,
ACCESSION			complete sequence.
KEYWORDS			AF163833 109116
SOURCE			AF163833.1 GI:5823209
ORGANISM			Streptococcus agalactiae
REFERENCE			Streptococcus agalactiae
AUTHORS			Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
TITLE			Streptococcus.
1 (bases 4755 to 6119)			
Rubens,C.E., Heggen,L.M., Haft,R.F. and Messels,M.R.			
Identification of cpsB, a gene essential for type III capsule			
expression in group B streptococci			
MoJ. Microbiol. 8 (5), 843-855 (1993)			
JOURNAL			
MEDLINE			
PUBMED			
8355611			
cpsB now called cpsE			
2 (bases 15471 to 16670)			
Haft,R.F., Messels,M.R., Mebane,M.F., Conaty,N. and Rubens,C.E.			
Characterization of cpsB and its product CMP-N-acetylneuraminic			
acid synthetase, a group B streptococcal enzyme that can function			
in K1 capsular polysaccharide biosynthesis in Escherichia coli			
MoJ. Microbiol. 19 (3), 555-563 (1996)			
JOURNAL			
MEDLINE			
PUBMED			
96228704			
8830246			
cpsF now called neuA			
3 (bases 1 to 17276)			
Chaffin,D.O., Beres,S.B., Yim,H.H. and Rubens,C.E.			
The serotype of type Ia and III group B streptococci is determined			
by the polymerase gene within the polycistronic capsule operon			
J. Bacteriol. 182 (16), 4466-4477 (2000)			
JOURNAL			
MEDLINE			
PUBMED			
20372630			
10913080			
4 (bases 2196 to 5903)			
Rubens,C.E.			
Direct Submission			
Submitted (02-JUN-1993) Division of Infectious Disease, Immunology			
and Rheumatology, Children's Hospital and Regional Medical Center,			
4800 Sand Point Way N.E., Seattle, WA 98105, USA			
5 (bases 4495 to 6194)			
Yim,H.H.			
Direct Submission			
Submitted (16-MAY-1996) Division of Infectious Disease, Immunology			
and Rheumatology, Children's Hospital and Regional Medical Center,			
4800 Sand Point Way N.E., Seattle, WA 98105, USA			
6 (bases 1 to 17276)			
Chaffin,D.O., Yim,H.H., Beres,S.B., Sweet,E.S., Niltayajarn,A. and			
Rubens,C.E.			
Direct Submission			
Submitted (29-JUN-1999) Division of Infectious Disease, Immunology			
and Rheumatology, Children's Hospital and Regional Medical Center,			
4800 Sand Point Way N.E., Seattle, WA 98105, USA			
On Sep 5, 1999 this sequence version replaced gi:1323729.			
COMMENT			
FEATURES			
Location/Qualifiers			





```

/product="capsular polysaccharide E"
/protein_id="BA033746.1"
/db_xref="GI:3721916"
/translation="MKIKLVSSGGHLMHNLKPKGNKDRFMTEDKASLIURE
EIVHCFPTNRNKNLVNNTLAFKTLREREDVLISSGAAVAFEFYIGKLGCKT
VYIEFPRIDKPLTGLVYPVTDKFIYQEEKKKVPKAINIGGIF"
2266..2808
/gene="cpsP"
2266..2808
/gene="cpsP"
2266..2808
/gene="cpsP"
/codon_start=1
/translation="galactosyltransferase"
/protein_id="BA033747.1"
/db_xref="GI:3721917"
/translation="WTFVTVGTHQOQNRKIKEDVRLKGTGALIDQEVFIQTGSDPEP
QNCQSKFSLYDDMSYKAEIVITHGPRATWNAVSKKNTIVPRQDQGEHVN
HODFLKELTLDYILNISELNTIKENISTSKVISONDDCFSEFNEHIVL
NKYLKKEIENISQISIC"
2847..3986
/gene="cpsG"
2847..3986
/gene="cpsG"
/codon_start=1
/translation="CP polymerase"
/protein_id="BA033748.1"
/db_xref="GI:3721918"
/translation="WQILLALIVLILCSYKQKMFYMAEPIFVPIYITLITL
LHSLFKTPDRIILAFNSLITIGVVALKRWNTLLELDKLIKALFNGLIFELG
GTYVYCHDNIQNIISIFGRDLISDMWNGHTQAFMGFEPSNLIIPMTVNTIYI
YMKLRNYSIMTIGVLLFTFTLPIYSGSGIYVAILQMFILNTVVAKKIKFLI
YILFPLVIMVLFNMLSIYRIINLRSSSESRSEYKIVTIVVNSLITGEEV
KELNLSPLTIGSHSTIYGYKSGGLIMNIVGGLIFINIGKKAQSAFYELIYG
TLITLPSFALEDIDGANMLIVIFLTVGLIENDFISQKRMKS"
3986..4987
/gene="cpsH"
3986..4987
/gene="cpsH"
/codon_start=1
/translation="N-acetylglucosaminyltransferase"
/protein_id="BA033749.1"
/db_xref="GI:3721919"
/translation="WEKRIIVSIIPIYNSAYKCEVQSYLQGTPLIIVILIDGS
TDSGELICDNLQEDNILLVFKHKGVSANMLDKSTGEPIFVDSDPVAPNMI
EIMLNLTENADIAVDFOISNEDYKCKKRAFYKFNNSLKEFLSGNRENIY
CTLYKKSIIIGNLFDEMLKIGEDLFRNCLCOEHRIVDTTSLYTRIVKTSAMK
OKTENSIIDPTITFNEVSLIVPAKLAVYKAKRERIKICLRKMFELGNSINDIKIKQ
REIFPKDKSPFKAVKYLKSLKSLFYLKCKSPKLYWAAVRFKIVABEIGENI"
4992..5939
/gene="cpsI"
4992..5939
/gene="cpsI"
/codon_start=1
/translation="galactosyltransferase"
/protein_id="BA033750.1"
/db_xref="GI:3721920"
/translation="MKIKIEDLISVPIYVVDYVVECIESIIVOTYNIETILIN
DGSITNCATIKAFSEBDCRIVLIEKSGISBANVGIYHSKVKLIFVDSODKSS
DYILNLVIAOKHDSIAIGLYEFERHNSIRNYELDKYIPEEALMLMDIKYQ
SIFITAMKLPKHSIFNDLEPALKYHEDRPNKAYLKAANSIYIDKPLVYHVRIG
SIMNSDNVILARKKLDVLSLDRIKLISLRKVSPLDKTEIFYNNOVPRTKKFLK
OQSVFKEDNTIDYRMTGRLRKVKLVDKLIKIRPF"
6023..6979
/gene="cpsJ"
6023..6979
/gene="cpsJ"
/codon_start=1
/translation="BA033751.1"
/protein_id="BA033751.1"
/db_xref="GI:3721921"

```

```

/translation="MNFVCHTLXHLITLIKDKENRIFICDTITDYETWKLIN
DQIRTSFKERAVRQLOMKYIEEWMELVSDLNHYPERVPTQVLEFDDTLIGYM
VYLGKRYHLITENGNCPOKFLGSSVYKRYIKYTLFKYVYSGSKVCLSLVNSLV
GLPHDIRSKQSLERFKRLPDSLNQKSLIKIRIKTRKPLTTPPSVLLTTPLODK
WKTKTPRQOSIOEYQFDDIVQSYRLGVNIVKTRHDVVDVDSKLEVEELSPNP
MELIEMSGRECOGITHSTALDLSLTDKRTITLVDKIDK"
6976..8376
/gene="cpsK"
6976..8376
/gene="cpsK"
/codon_start=1
/translation="capsular polysaccharide"
/protein_id="BA033752.1"
/db_xref="GI:3721922"
/translation="MNTKRLQSGFYTLGNLVQGLAFITLPIYTRVISTEVQGS
LYVANNITMLFICQTSLSAPVYKGEERKYSGSAFVGNIMFLLIILIAELFR
SELAIPVGSSEIFILMVCOSASYVTTRGOYFIQOQSLNLIISLNAVSVLS
LELIFRMSDPLSRVFGAVPTITGIVAFATITHSKSFNKRIFRIVTASVLEIF
HLHGQLQQLDRMLARLVNTRKAVMSRGSLSMI IQIVANSIMVAMI PWFEDARK
EKLDQSTYSRYLVGLVFLITGLVLPPELQIMGDRYSSVQFSLIIVSYELVF
LTFEPNIOFPVANTWIPIDITGLLAAGVAMNLVLI PHYAAYGAAMATII SYLALI
FHHIVSKRYHVSQSVRQYIILSGIVSYAMLMNMLDNVIVIRNELIILIVYSIV
EQVILDLISKRRRR"
8376..9401
/gene="neuB"
8376..9401
/gene="neuB"
/function="sialic acid synthesis"
/codon_start=1
/translation="sialic acid synthesis"
/protein_id="BA033753.1"
/db_xref="GI:3721923"
/translation="WYVILAEIGCNHNDINAKKXVDVAVSGVDVAKVQFRAKEL
ISKPAKAYOKETTGADSOLEMTNRLSEFSEYEMDYALSKEVETSTPDES
LEPLISTDMPYIKPSGRLTNLPYLERIKQKQKVLITSGMAVEIHQAVNLRQNG
TVDISLITCTEYTPPYPSLNTVHTLDEDFDILIGSDHSIGSEVPLAAMAGAE
VIRKFTLDTNMGSPDHKXATPDLIALYKGVRIYQALGRREKIPDPEERKNKIVA
RNSVALKPKTKGDIVSINETIVKRPNGISIPNMYDILIGQADDPDEDEVRDSRF
ENQLEPL"
9478..10206
/gene="neuC"
9478..10206
/gene="neuC"
/function="sialic acid synthesis"
/codon_start=1
/translation="sialic acid synthesis"
/protein_id="BA033754.1"
/db_xref="GI:3721924"
/translation="MKIKLVGSRABYGIKPELIQRLSKQSEVNTQIATAMHEEK
YGVYVQIREDGFDIAVKVPLHLYTDRETVSAPMAHLQGLKIKEDKEDYDVIILG
DRYEMLPVNVNALLIVPCHLHGSETSLANPEBYRHAATVMSHLIVSTEDFORV
IQNGRQPVINTGALGVENALSIPELTREATIRKQIGYIUBESFPVYLHPVTFEQQK
SAGBQKAVLALSKEFVQCLFIS"

```

ORIGIN

Query Match	30.7%	Score 324.8	DB 1	Length 10206
Best Local Similarity	58.2% <td>Pred. No. 2.1e-42</td> <td></td> <td></td>	Pred. No. 2.1e-42		
Matches 616	Conservative 0	Mismatches 427	Indels 15	Gaps 2

```

QY 1 ATGAGTAATATATATATCGTGTGCAAAATGTTGCAACCAATATGTTAGTGTGATAT 60
Db 8370 AAGATTAATGTTATATATATATGACAGATGTGTTGAATATATATGAGATATTAACCT 8429
QY 61 GCAAGGAATGATATTAAGCAAGAGAGCCGGTGTATATGACAGTAAATTCGAACA 120
Db 8430 GCGAAAAAATGTTAGATGTGCGGCTCTTGGGCTGTATGATCTGTAAATTTGACT 8489
QY 121 TTAAAGCTGATTAATTAATTTAGCTATTGCACTTAAGGCAAGATATCAATTAATAAAC 180
Db 8490 TTAAAGCTGAGAACTTAATTTCTAAATTTGCTCCCAAGCTGAATATCAAAAAAGAACT 8549
QY 181 AAGGAGAAATTAATCTGCTTGAAGATGCAAAAAAGCTTAATTAATGAAGTATGAGAT 240

```



[illegible]







[illegible]



Db 13791 GATCTAAGACCGAGTCGACACCTGATATTTTACGCTGCTTATGTTAAAGGGTTCGACT 13850  
 Qy 781 GTTGAAGAAATCTTTTAAAGTTCAAAATAGTAAAGTGTTCACGCTTCAGAAAGAAATTA 840  
 Db 13851 GTTGAAGAAATCTTTTAAAGTTCAAAATAGTAAAGTGTTCACGCTTCAGAAAGAAATTA 13910  
 Qy 841 ATCGTAGCAAGAAAGCTATTTAGCTTAAGACAGATATAAAAGGTAGGTTTTCATCA 900  
 Db 13911 ATTTGTTCTGCTTAATCACTGCTGTTTAAACCAATTAATAAGGCAATTTTATTTCA 13970  
 Qy 901 GAAAAAATATTAACAACAAAAGCTGTAAATGTATGATGCTGAGTAGAGTATAT 960  
 Db 13971 ATAGAAATATTAATGCTGAAGGCGCCAGGTAATGATATTCCTATGAACTGATATAC 14030  
 Qy 961 TTATTCGCTTAATTTGACAGACGACTTATTCAGATGATTAATATATATGACGA 1020  
 Db 14031 ATCTTGACACAAAGACGCAAGTATTCGAAAGAGATGAAGTTTGTGATTCAGCGC 14090  
 Qy 1021 TTCAAAATCAGGGGAAATATGAGAACAAAATTAAT 1058  
 Db 14091 TTTGAATATCAATTCGCCGAGTTATATACACTTAAT 14128

RESULT 10  
 AX602206 95596 bp DNA linear PAT 17-FEB-2003  
 LOCUS Sequence 135 from Patent WO02092818.  
 DEFINITION AX602206  
 ACCESSION AX602206  
 VERSION AX602206.1 GI:28402075  
 KEYWORDS  
 SOURCE Streptococcus agalactiae  
 ORGANISM Streptococcus agalactiae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.

REFERENCE  
 1 Glaeser, P., Ruzniok, C., Chevalier, F., Frangeul, L., Lailou, L., and Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Trieru-Cuot, P. and Kunst, F.  
 Streptococcus agalactiae genome sequence, use for developing vaccines, diagnostic tools, and for identifying therapeutic targets  
 Patent: WO 02092818-A 135 21-NOV-2002;  
 INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)  
 Location/Qualifiers  
 source 1. 95596  
 /organism="Streptococcus agalactiae"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1311"

ORIGIN  
 Query Match 30.7%; Score 324.8; Db 6; Length 95596;  
 Best Local Similarity 58.2%; Pred. No. 1.2e-42;  
 Matches 616; Conservative 0; Mismatches 427; Indels 15; Gaps 2;

1 ATGAGTAATATATATATCGTCTGTAATGTGTGCAACATATAGTAGTGTATAT 60  
 11392 AAGATATGTTTATATATATGAGAGATTTGTTGCAATCAATATGAGATATATCTT 11451  
 Qy 61 GCAGAGAAATGATATTAAGCCCAAGAGCGCGGTAAATGACATTAATTCGAACA 120  
 Db 11452 GCGAAAAAATGGTATATGTTGGGTGTCTTGTGGTGAATGCTGTTAAATTTGAGCT 11511  
 Qy 121 TTTAAAGTATTAATTAATTTCACTATGCACTTAAGGACAGATATCAATTAATAAC 180  
 Db 11512 TTTAAAGTATTAATTTCACTATGCACTTAAGGACAGATATCAATTAATAAC 11571  
 Qy 181 ACGAGAAATTAATTAATTTCACTATGCACTTAAGGACAGATATCAATTAATAAC 240  
 Db 11572 ACGAGAAATTAATTAATTTCACTATGCACTTAAGGACAGATATCAATTAATAAC 11631  
 Qy 241 TATTCGATCTAATGAAATGACAGCTTTTAAATTTGATGTTTCTACCCCTTT 300

Db 11632 TACTAGAAATGCGTATATATGCAATTTCAAAAGGTGAGACCTTTTCAACACTTTT 11691  
 Qy 301 GACGAAAGCTATATATATTTTATGCACTTTGAAACAAAATATAGAAATATCCCTTCA 360  
 Db 11692 GATTAAGATATATATATATTTTATTTTATGAGATATATGCAATTTTCAAAATTTCAATCA 11751  
 Qy 361 GGTAGTATATTAATTTTACCTATATCTTGAATAAATATGACCAAGCTTCCGATTAAG 420  
 Db 11752 GAGAAATCACTAATTTTACCTATATGAAAGATATGCAAGCAAC-----AAAG 11802  
 Qy 421 AAATATATATATCAAGAGAAATGCTATATGATATGATATGATATGATATGATATGAT 480  
 Db 11803 AAAGTATTTCTTGAAGCGGATGCGTATATGAAAGATATGCAAGCGGATATAT 11862  
 Qy 481 TTTATTAATTAATTAAGTTTCCGTTGTATATTAATTAATTAATTAATTAATTAATTAAT 540  
 Db 11863 TTACGTAGATATGATAC-----TGACATTTCTATTTTATGATATGATATGATATGAT 11916  
 Qy 541 CCAAGCGCTTTGAGATATTAACCTTAATGCTATTAATGATTTGAAATTAATTAATTAAT 600  
 Db 11917 CCAACCGCTTACCGCTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11976  
 Qy 601 AAGATTAATGATGCTCTGATCATCTTACGCGGTTTATGACGCTATTCGCGCGTG 660  
 Db 11977 GATTTAAGATATGATATATGATATGATATGATATGATATGATATGATATGATATGAT 12036  
 Qy 661 CTTATGAAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720  
 Db 12037 GCAATGGGTGCAAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12096  
 Qy 721 GATCATTTGGCTCAATGAAACCTGATGAAACCTGATGAAACCTGATGAAACCTGATGAA 780  
 Db 12097 GATCATTTGGCTCAATGAAACCTGATGAAACCTGATGAAACCTGATGAAACCTGATGAA 12156  
 Qy 781 GTTGAAGAAATCTTTTAAAGTTCAAAATAGTAAAGTGTTCACGCTTCAGAAAGAAATTA 840  
 Db 12157 GTTGAAGAAATCTTTTAAAGTTCAAAATAGTAAAGTGTTCACGCTTCAGAAAGAAATTA 12216  
 Qy 841 ATCTGAGAAAGATCTATTTATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900  
 Db 12217 ATTTGCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12276  
 Qy 901 GAAAAAATATTAACAACAAAAGCTGTATATGATATGATATGATATGATATGATATGATAT 960  
 Db 12277 ATAGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12336  
 Qy 961 TTATTTGTTAAATTTGACAGCAAGACTTTATTCAGATGATATTAATTAATTAATTAATTAAT 1020  
 Db 12337 ATCTTGACACAAAGACGCAAGATGATTTCCAGAGATGAAAGTATTTGATTCGATTCAGCGC 12396  
 Qy 1021 TTCAAAATCAGGGGAAATATGAGAACAAAATTAAT 1058  
 Db 12397 TTTGAATATCAATTCGCCGAGTTATATACACTTAAT 12434

RESULT 11  
 SAG766849 128050 bp DNA linear BCT 04-NOV-2002  
 LOCUS Streptococcus agalactiae NEM316 complete genome, segment 7.  
 DEFINITION AL766849 AL72656  
 ACCESSION AL766849 AL72656  
 VERSION AL766849.1 GI:23095613  
 KEYWORDS  
 SOURCE Streptococcus agalactiae NEM316  
 ORGANISM Streptococcus agalactiae NEM316  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.

REFERENCE  
 1 Glaeser, P., Ruzniok, C., Chevalier, F., Buchrieser, C., Frangeul, L., Zouine, M., Couve, E., Lailou, L., Wesdek, T., Poyart, C., Trieru-Cuot, P. and Kunst, F.  
 Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease  
 Mol. Microbiol. 45 (6), 1499-1513 (2002)

JOURNAL

MEDLINE 22242508  
 PUBMED 12354221  
 REFERENCE 2  
 AUTHORS Glaeser, P., Rusniok, C. and Frangeul, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-MAY-2002) Glaeser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86

## FEATURES

source

1. 128050  
 /organism="Streptococcus agalactiae NEM316"  
 /mol\_type="genomic DNA"  
 /strain="NEM316"  
 /serovar="serovar III"  
 /db\_xref="taxon:211110"  
 /complement(join(206..703,713..718))  
 /gene="gbs1139"  
 /complement(206..703)  
 /gene="gbs1139"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CAD46798.1"  
 /db\_xref="GI:23095614"  
 /db\_xref="SPTREMBL:Q8E593"  
 /translation="MSKLOQLIKEDPLRLTLNSPLHRSQTCDDYSIDTCDEVEQVYE  
 RLKGLIVLLKQQLKLSQMSRENRINCLTSYFETIAHRRASDSGDIAVLEVISI  
 MEDAAVGMANGIINNEFHAIICQMLNVEDYDKIEFKDKNFERFQOKKHGELF  
 EAFGL"  
 RBS complement(713..718)  
 /gene="gbs1139"  
 /complement(join(717..1469,1477..1483))  
 /gene="gbs1140"  
 /complement(717..1469)  
 /gene="gbs1140"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CAD46799.1"  
 /db\_xref="GI:23095615"  
 /db\_xref="SPTREMBL:Q8E592"  
 /translation="MNDILSELTKGLGYSNYSVNNMTAVNKSIAIGVILVLEFL  
 ELLSWYRFRNQGEWTKLFLVAVKVFAYLVASQALINAILMFLNVTYKLAGV  
 DLSGSLPEFKAKGAYGRIGLNFPLAVISIKWMLLRFILVYKAIAP  
 VIVAFMGDDTRIIANPLKTRGAILAQSAAVILILAIWQGFVDTRAMQLDNSWLGST  
 YAAISYIKCIEFLILFSCSKAKQLQVVS"  
 RBS complement(1477..1483)  
 /gene="gbs1140"  
 /complement(1495..2020)  
 /gene="gbs1141"  
 /complement(1495..1510)  
 /gene="gbs1141"  
 /complement(1518..2003)  
 /gene="gbs1141"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Unknown"  
 /protein\_id="CAD46800.1"  
 /db\_xref="GI:23095616"  
 /db\_xref="SPTREMBL:Q8E591"  
 /translation="MSESFYVAFADPFKRRKRPPTAELEKANVEGKDKSSRGER  
 LKGTLSFVIRNFRALLILQIPVYLTLPFNLKSAFAVAVICITIKAVPVITAEI  
 PFSQINDLSQAKPLIGFPAQDEMNNLSPIYFTRIDITICIFSVFALVMTFSKSR  
 V"  
 RBS complement(2015..2020)  
 /gene="gbs1141"  
 /complement(2109..2351)  
 /gene="gbs1142"  
 /complement(2109..2336)  
 /gene="gbs1142"  
 /codon\_start=1  
 /transl\_table=11

RBS complement(2346..2351)  
 /gene="gbs1142"  
 /complement(2611..5422)  
 /gene="gbs1143"  
 /complement(2611..5409)  
 /gene="gbs1143"  
 /note="putative peptidoglycan linked protein (LpxTG motif)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CAD46802.1"  
 /db\_xref="GI:23095618"  
 /db\_xref="SPTREMBL:Q8E589"  
 /translation="MNSQETKSGCFFRKSKAVGVCGIALAGATLTAISQVADQVTT  
 QATQTVQNAQAEVTSQDLKAVDTAKAAVAVTTTAVHATTTDAQADLANQVQ  
 VEDVYAKQAQNTQAIKDAIAEKAKIDAEKKEAEAEVAAANAGQAEVDAIRNAGQAAV  
 DARNAKQOQADQOKAKIDAEKASQSVQCLNQNKAKIDAEQDAQAKDATNAQCL  
 QEDVQTKLANIKSVAVAGVQRKQKQVADANAOLOQDYQAKALVQALAKA  
 EADKQSTNNVAFEDIKQAKGVQDNAEVQKSIPTAKTKPDGSEPPENIDMTDQIKTGVCK  
 LTGRKNHIVANKQSVTAFPDSVTLTKYETRYANAQNAANQNTYFRVLTDRGRPIFE  
 KAHNENKTPASTLNTLQNLKTBKPPHSSGNVSEFKIHDMVHDTGSALVSVNN  
 NDVAVNVAIPQPTPEKPEKTPERKKEVSPERKINVAHVQVQIPKATPTKVLVDENQSI  
 VPSIETPTPEKPTPEKPEKKEVSPERKINVAHVQVQIPKATPTKVLVDENQSI  
 NKSVLPNATLADVAKQNSQKQIKASAAIAKGFAPVQDQPNEMALAEIVYSIKASN  
 GDVASSILEMHHVLSKOTLDQKQSLIKAGISPVGEVYMTATKDPQAFQVAYQXGL  
 DITVNLSPKKKKEKFKQIQKNGVANDIDGNGTGNVITLTPDVPVHNDVLDKREKGS  
 INNGVYKLGDETYTLDEGVYVPAKNGYLPFEKPIVHQLHEDVLYKRYAKVAKATL  
 KDGVTPKGNLVQVTEFTVNTKETGRYLAERADFLAQVSSASAGADPTVYKAIKA  
 GDVNTADPFPVNGKVKTEVYVTPPEKPKVMPQKVPKAPALPSTGEQSVLTVL  
 GAALISLIGLVGFKRQO"  
 RBS complement(5417..5422)  
 /gene="gbs1143"  
 /complement(5476..6199)  
 /gene="gbs1144"  
 /complement(5476..6186)  
 /gene="gbs1144"  
 /note="putative peptidoglycan linked protein (LpxTG motif)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CAD46803.1"  
 /db\_xref="GI:23095619"  
 /db\_xref="SPTREMBL:Q8E588"  
 /translation="MNRKRLTVLSIAVAVLSVTPADEIPSPDPAFSPPLINPTP  
 VDQGPPTTKKSPVPDQGPPTDTPKSTPVDPQGPPTTKKSTPVDSSTNPSDSB  
 AITNMIGTKRPDQVYKPTPKIPKPSVTPPVQIDVDEKNGTITIKTIDSPDKTIIG  
 TQNGVITRDSGSHVASSEVGKRLNDGVYITIKDSSGKTKTLPTRTGAKGFLSIIG  
 AILISFVAVLPKQVTLN"  
 RBS complement(6194..6199)  
 /gene="gbs1144"  
 /complement(6201..8447)  
 /gene="gbs1145"  
 /complement(6201..8432)  
 /gene="gbs1145"  
 /note="similar to surface exclusion protein Sec10 precursor - Enterococcus faecalis plasmid pcef10 (LpxTG motif)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CAD46804.1"  
 /db\_xref="GI:23095620"  
 /db\_xref="SPTREMBL:Q8E587"  
 /translation="MNCIKITIGLVATLSAVGVNVAEDITTPAPVNEPOVSESTAK  
 TPQVTEQVNSAKVTAQDQATSDVNAQKVVNNAQKQKQAOQKLVNATTTINETOKLV

QESTNONGVOQTVDASAKQRLSQTETANOKITQTEQVAKONOVNAQQIVAVNNEHDVATK  
TADYKQASVYDIAKDALNTNTI VNSDLKQASNTTQADYKATIDALITRAQADKTL  
TNQAKQOIVDSAKQNSLADTOLQSNANVNHKEFTALGQSHYNOQADYK  
GGEFTASTGCVPSALAMWYSDLSNRTIPEPILADYLVNNTDEPKRFGSGISKIIISA  
TKAFGVYVTHIASKNATLREALKGHVVAVAVONKFSFMPQVSHIIVLGGSSGNTY  
VYDPYNDNNGFYSYDRIHNEGSDSIDTAGVCPFPALIMTKMANALITQSOALASO  
QVAKQONDAQAKATGNAVTOTPTIAGANTIKASNTKDKQKRLAARQASVKLAND  
NVKQADITKASBKLDKQKQALAAQALITSKTLNOLKQVLAESQVAAQADYK  
QAKNLTQKLVYLTMLRNAQNLIRKQSDVAAQADNLANKIAKLQREVALYBELTKA  
VDAQSOYKVLKAVKSVLSAKASLKLAEKARLDKREAVAVAVDETKITSYTSKHE  
KLEKSLVATKTTDVKQVSAKASVLPSTQDVQVVALJLMLLTFSGFLGRQSKK  
VIN"

complement (8442.. 8447)  
/gene="gbs1145"  
complement (8449.. 8759)  
/gene="gbs1146"  
complement (8449.. 8748)  
/gene="gbs1146"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CAD46805.1"  
/db\_xref="GI:23095621"  
/db\_xref="SPTREMBL:Q8E586"  
/translation="WDETKKYLRLDLNLKQGGATRLDEGTELFLKPDYAVRTAK  
GYVQIPIPDVEFHLTYRSIFNQIRTIKKQGHVAKERSPSGLVQLTGKGYKRP"  
complement (8754.. 8759)  
/gene="gbs1146"  
complement (9131.. 9328)  
/gene="gbs1147"  
complement (9131.. 9328)  
/gene="gbs1147"  
/note="Similar to phage protein"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CAD46806.1"  
/db\_xref="GI:23095622"

Query Match 30.7%; Score 324.8; DB 1; Length 128050;  
Best Local Similarity 58.2%; Pred. No. 1.1e-42;  
Matches 616; Conservative 0; Mismatches 427; Indels 15; Gaps 2;

1 ATGAGTAATATATATATGCTGCTGAAATGCTGCAACCATATGTAAGTGTGATTT 60  
90254 AAGATATATGCTTATATATATGCAAGATGCTTCAATCATATATGAGATATATATCTT 90195  
QY 61 GCAAGGAAATGATATTAAGCCCAAGAGCGCGGTGTAAGCAGTAAATTCACAAACA 120  
DB 90194 GCGAAAAAATGATGATGCTGCGTCTTGTGTGTGATGCTGTTAAATTTTCAGACT 90135  
QY 121 TTTAAAGCTGATATATATTTTCACTATGCACTTAAGCGAGAGATATCAATTAATAAAC 180  
DB 90134 TTTAAAGCTGAGAACTTATTTCTTAATTTGCTCCCAAGCGAATATCAAAAAAAGAACT 90075  
QY 181 AAGAGGAAATGATCTGATTAAGTAAGCAAAAAAGCTTGAATGAAGTATGACAT 240  
DB 90074 AAGAGGAAAGCGACAGTCACTTAAGTGAAGCAAAAGCTTGAAGTAAAGCTTTGAAGAA 90015  
QY 241 TATCTCATCTATGATATGATATGCACTGATTAAATTTAGATGTTTTTTTCAACCCCTTTT 300  
DB 90014 TACTTAGAAATGCGATATATGCAATTTCAAAAGGTGAGACCTTTTCAACACCTTTT 89955  
QY 301 GACGAGAGCTTATGATTTTTTACGATCTTTGAACAACAAAAATATGAAAAATCCCTCA 360  
DB 89954 GATGAGAGATCTATGAGTCTTAAATTTTACAGATATGCCAATTTTACAAATTTCCATCA 89895  
QY 361 GGTGAGATTTGAAATTTACCGTATCTGAAAAAATAGCAAGCTTCCGATCCCTGATAG 420  
DB 89894 GAGGAAATGATATTTTACCTTACTTAGAAAAATGGCAACCAAC-----AAAG 89844  
QY 421 AAAATATATATCAACAGAAATGCTACTATGATGAGATTAACAGCTCTTTCTATT 480

DB 89843 AAAGTATTTCTTTCGACGGGTATGCGGTATATGGAAGATCATCAAGCGGTGAATAT 89784  
QY 481 TTTATATATATATATAGTTCGGTGTGATATTTTACATATTTACATTTGCAATATGAAATAT 540  
DB 89783 TTAGCTGAGATATGTAAC-----TGACATTTCTATTTTATATGTAACAGAGTAC 89730  
QY 541 CCAAGCGCTTTGAGATATGAAACCTTAAGCTATTAAGATTTGAAAAAACAATTCCT 600  
DB 89729 CCAAGCGCTTTACCCCTCTTAAATTTTAAAGTTTATCTTATTTGAAAGATTTTAA 89670  
QY 601 AAGATATATAGCTCTCTGATATCTTACAGCGGGTTTATGACGATATGCGCGCTG 660  
DB 89669 GATTTACGATATGATTTTTCGATATCTCAATTTGATCAAGATGACTTATGACGACGA 89610  
QY 661 CTTATGGAATATCTTTTATTTGAAAAACATTTGACTTATGATATATCTATGCTGCGCA 720  
DB 89609 GCAATGGGTGACAGATATTTGAAAAACCTTTACTTTAGATATATGAGAGTCCG 89550  
QY 721 GATCATTTGGCTCAATATGACCTGATGAATGAAATGAAATCTTTGATTTGAGGTGAGGT 780  
DB 89549 GATCATATAGCGAGTGCACACCTGATATTTTATGCTGCTTATGATTAAGGCGGTGCAAT 89490  
QY 781 GTTCAAAATATCTTTAGTTGAAATATGAAAGTGTATACACTTCAGAAAGAGATATTA 840  
DB 89489 GTTGAACAGCTTATGATATTTGAAAAATCCAGATCCAGTACAGAAAAATATAG 89430  
QY 841 ATCGTACAGAAAGTCTATATATAGCTTAAACAGAGATTAATAAAGGTGAGTTTTC 900  
DB 89429 ATTTGCTGTGTAATATGAGTGTGTTTAAACATTAATAAAGCGCATATTTATTTCA 89370  
QY 901 GAAAAAATATATACAAACAAAAGCTGTATATGTTACGTGCGAGAGGTGTATAT 960  
DB 89369 ATACAAATATATATGCGTGAAGCGCCAGTAAAGTATTTCTCTAGAACTGTATGAC 89310  
QY 961 TTTATGCTTAAATATGCAAGCAAGACTTATTTCCAGATATTAATATATATATGCGAA 1020  
DB 89309 ATCTTGACAGCAAGAGCGCAAGATTTTGAAGAGATATGATTTGATTTGACG 89250  
QY 1021 TTTCAAAATATGAGGAGATATATGAAACAAATATAT 1058  
DB 89249 TTTGAAATATATGCCCCGATATATATACACTTAT 89212

RESULT 12  
AK607023 1026 bp DNA linear PAT 17-FEB-2003  
LOCUS Sequence 4952 from Patent WO02092818.  
DEFINITION AK607023  
ACCESSION AK607023  
VERSION AK607023.1 GI:28402552  
KEYWORDS  
SOURCE Streptococcus agalactiae  
ORGANISM Streptococcus agalactiae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE  
1 Glaser, P., Rusniok, C., Chevalier, F., Frangeul, L., Lailou, L.,  
Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Tieu-Quot, P. and  
Kunst, F.  
Streptococcus agalactiae genome sequence, use for developing  
vaccines, diagnostic tools, and for identifying therapeutic targets  
Patent: WO 02092818-A 4952 21-NOV-2002;  
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE  
SCIENTIFIQUE (CNRS) (FR)  
FEATURES  
source  
1..1026  
Location/Qualifiers  
/organism="Streptococcus agalactiae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1311"

Query Match 30.3%; Score 321.4; DB 6; Length 1026;  
Best Local Similarity 58.8%; Pred. No. 1.3e-41;  
Matches 600; Conservative 0; Mismatches 466; Indels 15; Gaps 2;



```

/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21599.1"
/db_xref="GI:12697596"
/translation="M1PVTVGTHGQPRNLIKEDVRLKGTGADIDGVPIQGYSPDF
QNCQMSKPLSYDDNMSYMKAEIVTHGGPAPFNAAVSKKKTIVAPRQDQGEHVN
HGVDFLKEKLFLEKELDYILNISELNIKEKNIISTSKVLSQNDPCSSFKQELSKLFE
"
RBS
2806..2812
2842..3981
/gene="cpsIBK"
2842..3981
/gene="cpsIBK"
/function="putative CP polymerase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21600.1"
/db_xref="GI:12697597"
/translation="MQLLLALLVLLCCSYNEKPKFLNNAIFFIYVYVYVSVISV
LNSLFRSPFHFVIAAFNSIAGVWSLFFHYKNTNIELTLKSLFNALILFCIG
FLYYAIVDFVENSLFGRLIGSDINGMTHORAMAFPEYSLIPLIINIIYIYI
YIKORSSGMMLIGALSTIILPISGSGAGIIVALLVLIILLNIIYIKQOTIEFL
YVPIILILVILRPNLWVSIYRNINLPSGSESPSLKQDYSVITDSTFLGKV
KELWMSDLPLSGSHSTYIGTFKIGLPGILNITLIGFLILISITIKAAKSDPYEIVG
SVLLFSFPALEDIGANWLIIFVTVGLIKNKDPSQLKRWES"
3964..3970
/gene="cpsIBK"
3961..4955
/gene="cpsIBK"
3961..4955
/gene="cpsIBK"
/function="beta-1,3-N-acetylglucosaminyltransferase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21601.1"
/db_xref="GI:12697598"
/translation="MEKQILVLSVITPYNSEAYLKACVQSVLQOTHSLEVLINDGS
TNSGELCDNLSQKDRILVFKKNGVSSARLGLDKSTGEITTVDSDDVAPLIT
EIMLKNLIEDADIAVDPDISNERDYRKKRNFYKFNKNSLSKEFISGRVENVIT
CMLYKSLIIGNRFDENLKIGEDLFNCILQEHCIIVDTISLYTRIVYTSAMN
OEENENSIDRTITFNISISIVPAKLANYEAKFLRKKRCIARZMPELGSNIDSKILO
RATPFKDYKIPFYKAVKYSLSKGLSIYLMKCSPLIYIKLDRPKQK"
4995..5001
5008..5949
/gene="cpsIBK"
5008..5949
/gene="cpsIBK"
/function="beta-1,3-galactosyltransferase"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21602.1"
/db_xref="GI:12697599"
/translation="MNYSTMSYNEPYNVARDVSISLNOULTDEFFIIVINDPSRG
DKQPLREVSVDNRKIILNERNIGLASSLNKAVSISGEYIFRMDADISPSFD
KQIRFENSLDPSATILILIDOKNIVYKORSNKIYILNDRKLLRSILIAHTW
KQKRYVADKMGFRDLPVADYPAIRGALADFKIGLKNVLLQVLTANLNGISQTKPF
CXXKYSAILQDPFYKESYIDITKIIVPQSYVKKRYTQDELKSYELKSTSTITRK
LYICLYYFKSPVLRLLINDINILVRLKFGKQSD"
5928..5935
/gene="cpsIBK"
5942..6880
/gene="cpsIBK"
5942..6880
/gene="cpsIBK"
/function="putative sialyltransferase"
/codon_start=1

```

```

/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21603.1"
/db_xref="GI:12697600"
/translation="MTNRKLIYVCHTLVHLICLYKEEISYNLIELIISSSIPDVNLEK
KLKSTNIIHLERSGSEBSLELVKDGLSYKPSNCFIENDATPIGRLLIKHGI
YTNILNGLNCFTYSISQCLMKYKVKYILHKKIOPGSPRCVLCITQPSLIDKESYTN
RYKPIEIVPRKELNDVTEYOKEMALNLPKAVRSTISPSVLCITQPSLIDKESYTN
NKIETSEQNFYKSIIVNEYINKGVNLYLVHPDVVDYSKLPVELLPSVPMETIEL
MLNRFEGGITHSETALDFLCVDRKILTVLDRKIK"
6865..6870
/gene="cpsIBK"
6877..8277
/gene="cpsIBK"
6877..8277
/gene="cpsIBK"
/function="putative repeat unit transporter"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21604.1"
/db_xref="GI:12697601"
/translation="MTNRKLLQSGRIYTLGNLAVQGLAFTLPIYTRVISVEYQGS
LYVAMNIMLPIGQTSGLSARVKYGEFEKYSGSAPSVGNIMPLIALLFLR
SLAPLVGSPSIFLWVCOSSYANVVTFFPQYFLOQORSLANILSLANVSSVALS
LPLFPHSDPDLSEVLGVFPTLITGVAPAYIYHSSFPNPKYFRIIVTSYLLIF
HLHQLNQLDRIMELNLTSTKRYAMSRSYSGMIIQIYLSINMAMIPMPDARK
ELIQLSTYISRYVLYGVFLGVTVFPELQIMGDDQSSVQPSLITIVSYTARV
LYTLPVNIQFPFYNANTMNPICGLLAAGVNLNVLPHVAYGAMATISYLLALI
FHHVSKVYKHYSDVSRYIILSGIYFVSAMNMFLNIVISWSGLIILIVYSIV
FOKYLIDLSKKRRR"
8262..8269
/gene="cpsIBK"
8277..9302
/gene="neut"
8277..9302
/gene="neut"
/function="sialic acid synthetase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="neut"
/protein_id="BAB21605.1"
/db_xref="GI:12697602"
/translation="WVYIIAIEIGNHNNDINLAKKMDVAVSCGVDAVKFQTEAKXL
ISKAPKAYOKATGTADSOLEMTKRLSEFEELYEMRDYALISKGVTFSTFPEDES
LRLISTGMPYIKIPSGRIITNLPLYEKIKGQOKVYLSTGMAVMEIHQAVNIIQNG
TTDISIHCCTBYRPPYPSINLNTHTLKKRKHDTIGYSDHSISGEVPIAAAGAAE
VIEKFTPLDNNRVEDHRSASNPDIILALVKGFPALINQALGRFEKIPDPVEKKNIVA
RSVVALPKIKKGDYISLENITVRRPANGISFPMWYDILGDAQDDFDEVDITDRSR
ENQLEBEL"
9367..9374
9379..9987
/gene="neut"
9379..9987

```

Query Match 29.5%; Score 312; DB 1; Length 9987;  
Best Local Similarity 57.5%; Pred. No. 2.2e-40;  
Matches 608; Conservative 0; Mismatches 435; Indels 15; Gaps 2;

```

1  ATGAGTAATATATATATGCTGTAATATGTCGCAACCAATAATGATGATAT 60
Db 8271  AAGATATATGTTATATTTATTCAGAGATGCTTGCATCAATAAGAGATATATCT 8330
QY 61  GCAAGAGAAATGATATTAACCCAAAGAGCCCGGTGTAATGCGTAAATTCCAACA 120
Db 8331  GCAAAAAAATGATAGATGTCGCCGCTTGTGCGTGTGAGACGCTGTAATTTAGACT 8390
QY 121  TTTAAGCTGATTAATTAATTCAGCTATGACCAAGGCGAGATCAATTAATAAAC 180
Db 8391  TTTAAGCTGAGAACTTATTTCTAATTTGCTCCCAAGCTGATATATCAATAAGAACT 8450

```

	Y	181	ACGGGAATATTAGAACTCAGTTAGTAATAATGCAAAAAAGCTGAAAATGAAGTATGACAT	240
	D	8451	ACGGGAACAGCACAGCTGCACACTTGAGATGACGAAAGCTTAGAGTTAAGCTTTGAAGA	8510
	Q	241	TATCTCCATCATATGGAATAATGACGTAGCATTTAAATTTAATGATGTTTTCTACCCTTT	300
	D	8511	TACTTAAANAATGCGGTATATGCAATTTCAAAGAGTGGAGACCTTTTCAMACCTTTT	8570
	Q	301	GACGAACACTATATGATTTTATAGCATCTTTGAAACAAAATAATGAAATGCCCTCA	360
	D	8571	GATGAAGAAGTCATTAGAGTTCTTAATTTCTACAGACATGCCAATTTCAAAAATTCATCA	8630
	Q	361	GGTAGCTTATGAAATTTACCGGTATCTGAAAAAATAGCCAAAGCTTCCGATCCGTATAG	420
	D	8631	GGAGAATTCACCTAATTTACCTTACTTGAAAAAGATTGGCAAGCAAC-----AAAA	8681
	Q	421	AAATATCATATCAACGAATATGGCTATCTATATGATGATAAATAAGCTCGTTCTATT	480
	D	8682	AAAGTATCTTTGGAOOGGTATGCGGTATAGGAAGATCCATCAAGCGGTAAATTT	8741
	Q	481	TTTATTAATATAAATGAATGCCGGTGGTAATTTTCAATATTACATTCGCAATCTGAAT	540
	D	8742	TTAGTCAGATGATGACACCG-----ACATTTCTAATTTTACATTTGACAACAGATAC	8795
	Q	541	CCAAGCCCTTTGAGAGATTAACCTTAATCTATATGATTTGAAAAAACATTCCTCCT	600
	D	8796	CCGACACCTTACCCCTCTTAATTTAAAGTATATCAATCTTGGAAAAATGAATTTTAA	8855
	Q	601	AAGATAATCATAGGCTCTCTGATCATCTCTCCGCGGTTTATGACGATATTCGCGCGTG	660
	D	8856	GATTTAOCGATAGGTATGCGATTCATTCATTCGATGGATCAGAAATACCTATCGACAGCA	8915
	Q	661	CCTATGGAATTAATCTTTATGAAAAACATTTCACTTTAGATAATCTATGCTCGGCCA	720
	D	8916	GCAATGGGTGCAGAGATTTAGAAAAACACTTCTCTTGAATCTAATATGGAAGTTCCG	8975
	Q	721	GATCATTTGGCCTCAATAGAACCTGATGACCTGAACATCTTTATTTGGGTCAGGTGT	780
	D	8976	GATCATAAAGCGACGTGCAACACCCGATATTTAGCTGCTTATGTTAAAGGTTGCAATG	9035
	Q	781	GTTGAAAAATCTTTTAGTTCAATATATGTAATGTAAGTGTACGTTACGTTAGAAAGACAAATPAA	840
	D	9036	TGGAACCAAGCCTTAGGTAGTTGAAAAAATCCAGATCCAGTGAAGAGAAAAAATPAG	9095
	Q	841	ATCGTAGCAAAAAGTCTATATATAGCTTAAACAGAGATPAAAAAAAGGTGAGTTTTTCA	900
	D	9096	ATTGTGTCTGTAAATCAGTCGCTTGCTTTAAACCAATTAATAAAGCGAATTTATTTCA	9155
	Q	901	GAATAAATATPACACAAAAGACCTGTGTAATGATCACTCCGATGAGTGTATAT	960
	D	9156	ATAGAAAAATATVACCGTGACGCCCAAGTATGTGATTTCTCTATGATACCTGTATGAC	9215
	Q	961	TEATTGGGTAAATATGACAGACAGACTTTATCCAGATGATTAATATATCATAGCGAA	1020
	D	9216	ATCTTGGACAAAGAGCCAGATATATTTCCAGAGGATATGAAGTATTCGTATTCAGCG	9275
	Q	1021	TTCAAAAATCAGGGGGAATTAATGACACAAAATTAATT	1058
	D	9276	TTTGAAATCAATGCCCCAGATTATTAATACCTTAAT	9313
<hr/>				
RESULT_14				
LOCUS	CJ11168X5	263335 bp	DNA	linear BCT 09-JUL-2002
DEFINITION	Campylobacter jejuni subsp. jejuni NCTC 11168 complete genome;			
ACCESSION	AL139078 AL111168			
VERSION	ALI39078.2 GI:6968723			
KEYWORDS				
SOURCE	Campylobacter jejuni subsp. jejuni NCTC 11168			
ORGANISM	Campylobacter jejuni subsp. jejuni NCTC 11168			
	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;			

```

REFERENCE
AUTHORS
Campylobacteriaceae; Campylobacter.

1
Parkhill, J., Wren, B.W., Mungall, K., Kelsey, J.M., Churcher, C.,
Besham, D., Chillingworth, T., Davies, R.M., Meltwell, T., Holroyd, S.,
Jagels, K., Kariyasekera, A., Moule, S., Pallen, M.J., Penn, C.W.,
Quail, M., Rajandream, M.A., Rutherford, K.W., Vanniel, A.,
Whitehead, S. and Barrell, B.G.
The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences
Nature 403 (6770), 665-666 (2000)

JOURNAL
MEDLINE
20150812
PUBMED
10686204
REFERENCE
2 (Bases 1 to 263335)
PARKHILL, J.
DIRECT SUBMISSION
Submitted 109-FEB-2000) Submitted on behalf of the Campylobacter
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENT
Details of C. jejuni sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/C\_jejuni/).

FEATURES
source
1. 263335
location/Qualifiers
/organism="Campylobacter jejuni subsp. jejuni NCTC 11168"
/mol_type="genomic DNA"
/strain="NCTC 11168"
/db_xref="taxon:192222"
complement (38..1369)
/gene="accC"
complement (38..1369)
/gene="accC"
complement (38..1369)
/gene="accC"
/EC_number="6.3.4.12"
/note="Cj1290C, accC, probable biotin carboxylase (subunit
of acetyl-CoA carboxylase (EC 6.4.1.2)), len: 443 aa;
highly similar to many e.g. ACCC EC01 biotin carboxylase
(BC 6.3.4.14) (449 aa), fasta scores: opt: 1560 z-score:
1741.0 E(): 0.35.2% identity in 442 aa overlap, 67.0%
identity to Hp0370. Also similar to Cj1037c pycA (46.5%
identity in 445 aa overlap). Contains PS00866 and PS00867
Carbamoyl-phosphate synthase subdomain signatures 1 and 2,
and Pfam match to entry PF00289 CPSase_L_chain,
Carbamoyl-phosphate synthase (CPSase)_"_chain.
/codon_start=1
/transl_table=1
/product="biotin carboxylase"
/protein_id="CAB73717.1"
/db_xref="GI:6968724"
/db_xref="GOA:O9PN09"
/db_xref="SPTRMBL:O9PN09"
/translation="METIKSLIINRGIALRAITIKEMGKKAICVSEADKALYLK
YADSLGICGKASSESYLNIPIATIAAEIABDAIPFGGSLASNOMFPERCAGNLIK
FTIGSVANIMSDRSKAKYOMQAGPVAIPGSGALAGAAKAKLKEISYPIILKR
AAGGGGKGVENEMKDLERAKYMSAESBAMTAFEDGTMYEKYIQNPHIEVQYIGSS
FGNVIHGEEDSCMORRHKLIIESPAIILDEKRTILHETAIIPAALIGYAGTSEE
FLVNDLDFEIKETILQVHECVSEVSGIDILEQMIKVAEGVALPSQSEIKINGHSE
IECHITAEDEKTEFLPSGKITKYIPAPGAVRWASHCGQDVSAYVDSMTGKLVMVA
EDRKAIARAKVMDRLISICIGIKTTDHFILSMENPDPIINNNTYTLAR"
complement (47..1357)
/gene="accC"
/note="Pfam match to entry PF00289 CPSase_L_chain,
Carbamoyl-phosphate synthase (CPSase), score 640.70,
E-value 8e-189"
complement (488..511)
/gene="accC"
/note="PS00867 Carbamoyl-phosphate synthase subdomain
signature 2"
complement (866..910)
/gene="accC"
/note="PS00866 Carbamoyl-phosphate synthase subdomain
signature 1"
complement (1371..1826)

```



```

CDS
/gene="accB"
/complement(1371..1826)
/BC_number="6.4.1.2"
/notes="Cj1293c, accB, probable biotin carboxyl carrier protein of acetyl-CoA carboxylase, len: 151 aa; similar to e.g. BCCP-ECOLI biotin carboxyl carrier protein of acetyl-CoA carboxylase (EC 6.4.1.2) (156 aa), fasta scores: opt: 307 z-score: 329.6 E(): 4.6e-11, 38.9% identity in 162 aa overlap, 41.4% identity to HP0371. Contains Pfam match to entry PF00364 biotin_req_enu, biotin-requiring enzymes"
/codon_start=1
/transl_table=11
/product="putative biotin carboxyl carrier protein of acetyl-CoA carboxylase"
/protein_id="CA873718.1"
/db_xref="GI:6968725"
/db_xref="GOA:Q9PN08"
/db_xref="SPTREMBL:Q9PN08"
/translation="MTKEIKELVNLFAEANTSKIKIKRQGFETLEFDMCCVPAPVCPVAPQPIVNVNEAPQSVSNRPSINSPMWTFYQAPSPGAPAFKVGSTVKKGDPIATIEAKKIMNEIABDCRIASILVADQPVFEGMPLFAVERL"
/complement(1383..1592)
/gene="accB"
/notes="Pfam match to entry PF00364 biotin_req_enu, biotin-requiring enzymes, score 104.20, E-value 2.5e-27"
/codon_start=1
/transl_table=11
/product="putative deoxycytidine triphosphate deaminase"
/protein_id="CA873719.1"
/db_xref="GI:6968726"
/db_xref="GOA:Q9PN07"
/db_xref="SPTREMBL:Q9PN07"
/translation="MGTKADNMTRKMALEKHMIEPCENATIGKGVSYGSLGYDIRVVERKIFTNVNSTVDPKPFVEENVVDREGVCIYRANSPLATITEYRKPENVTALICLGGSTVARCGIVNVTPEFEGSHITETISNTTLPRAKIYANEGIAQVLFQSDSKCDTLYKDKKRGXQAQTGITLPRILK"
/2046..2522
/gene="dcd"
/notes="Pfam match to entry PF00692 dUTPase, dUTPase, score 3.80, E-value 0.0016"
/2584..3588
/gene="Cj1293"
/2584..3588
/gene="Cj1293"
/notes="Cj1293, possible sugar nucleotide epimerase/dehydratase, len: 334 aa; similar to many e.g. TR-Q45984 (EMBL:U27301) Caulobacter crescentus FLA1. Protein (331 aa), fasta scores: opt: 1240 z-score: 1410.9 E(): 0, 57.8% identity in 325 aa overlap, CAPD_STRAND CAPD protein (559 aa), fasta scores: opt: 529 z-score: 602.6 E(): 2.9e-26, 33.9% identity in 319 aa overlap, and TR:069130 (EMBL:AF064070) Burkholderia pseudomallei putative epimerase/dehydratase WBII (637 aa), fasta scores: opt: 433 z-score: 493.6 E(): 3.4e-20, 30.5% identity in 282 aa overlap, 64.6% identity to HP0840. Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="possible sugar nucleotide epimerase/dehydratase"

```

```

/misc_feature
/protein_id="CA873720.1"
/db_xref="GI:6968727"
/db_xref="GOA:Q9PN06"
/db_xref="SPTREMBL:Q9PN06"
/translation="MPEKKNILLTGGSFGKTYKVLLENYXNKIILYSRDLKOPEMASVPAPGMYIGIVDRKERSAMRQVDFVTHAAKQHYIAENMECIKTIHGMNVYDAPFNVQVCCIALSTDKACNPVINGATLADKLFEVANNAGKOTFQVTRGVNVSSRGSVDPFKKLISEKKEPIPIIDPTWTRWISLQGVKLVNEERMHGSEIFPKIPSKMTKIDTALADLANVSHKILIGRAGKHEIMNISDSDSHLYTERFYVAISPIKPIFDKNDPFSINALGEKQKXVDSNNPLMASEKSLLEIINTEBP"
/2617..2640
/gene="Cj1293"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
/3590..4720
/gene="Cj1294"
/3590..4720
/gene="Cj1294"
/notes="Cj1294, probable aminotransferase, len: 376 aa; similar to many members of the degT family e.g. SPEC_BACSU spore coat polysaccharide biosynthesis protein (389 aa), fasta scores: opt: 586 z-score: 676.7 E(): 2.1e-30, 35.3% identity in 385 aa overlap, and TR:088001 (EMBL:AJ007747) Bordetella bronchiseptica putative amino-sugar biosynthesis protein WBC (366 aa), fasta scores: opt: 482 z-score: 558.1 E(): 8.7e-24, 28.8% identity in 330 aa overlap, 44.8% identity to HP0366. Contains Pfam match to entry PF01041 DegT_Dnrd_EryC1, DegT/Dnrd/EryC1/Stx family"
/codon_start=1
/transl_table=11
/product="putative aminotransferase (degT family)"
/protein_id="CA873721.1"
/db_xref="GI:6968728"
/db_xref="SPTREMBL:Q9PN05"
/translation="MTYSRQNTQSGSIDITKALKDLEILGCKXNVEFELCEWGVKRAVCNARSALSHLAVYTAGVQEKVLTLPFLPATAANLMAKQVETIDKNGNIDERKLEARLKSSENIGALSVDYFGNSXEMDEINLTKKVIPLIDDSHLAGALDYSEKVGKKADLSIFSPHVPKPIITTFEGGAVSNBELIDIKILRSHGIYKXLMDSIMVELGVNYSRLDVAALGINOLKDLHNEKREINLNFYKPEKPKYPTSTIKIYKKSSRHLPIILFPFVQKKEELFESLHAGIQVYKPFYFESFYKLLGEIKLQYADNRKAFSLIPCHQENLMDAKRVKDTLPISLEKVKGYGC"
/3608..4681
/gene="Cj1294"
/notes="Pfam match to entry PF01041 DegT_Dnrd_EryC1, DegT/Dnrd/EryC1, score 300.6; DB 1; Length 26335; Best local similarity: 57.7%; Pred. No. 6,4e-19; Matches 581; Conservative 0; Mismatches 414; Indels 12; Gaps 2; Query Match 1
1 ATGAGTAATATATATATGCTGGAATGGTGCACCAACCTAATGCTAGTGATTT 60
31867 ATGAAAAAACTTATATATGACAGACAGAGATTAATCAATGGGATTTAAATTA 31926
61 GCAAGAGAAATGATATTAAAGCCAAAGAGCGGGTGAATGACGTAATTCACAACA 120
31927 GCTAAAGAGTTAATGAAATGACAGACACTTCAGAGCTGATTTGTTAATTTCAAGT 31986
121 TTTAAGCTGATTAATTAATTTACGTTATTCACCTTAAGGACAGATCAATTAAC 180
31987 TTTAAGCCAAAACTGATACGACAAAGGCTTAAGACCTTACAGCTTAAACG 32046
181 ACAGAGAAATGAAATTCAGTTGAATGACAAAGAGCTTGAATGAAGATGACAT 240
32047 ACGGTAATGATGAAGTCAAGCTTCAATGTGCAAAAACCTGAACTGACTTAAAGT 32106
241 TATCTCATCTAATGATATGACAGTCAATTAATTTAGATGTTTCTTACCCCTTT 300
32107 CATAAAGGCTTATCTTACGCAAAAAGCAATATGCGCTTCTTCTTACCTCTTT 32166
301 GACGAGAGCTTATGATTTTTCAGCACTTGAACAAAGAAATATGAAAAATCCCTCA 360
32167 GATCTGAAGGTAGATCTTTTAAATGATGAGCTTAAAAATCTTTAATATTCAGAC 32226

```

QY 361 GGTCATTATTGATTTACCGTATCTTGAAATAATGAGCCAGCTCCGATCCCTGATAG 420  
 DB 32227 GGAGAAATGACCAATCTACCTCTATCTTAAATAATGCAAGCT-----TATATA 32277  
 QY 421 AAAATATCATATCAACAGGAATGCTACATTTGATGATGATTAACAGCTGTCTTAT 480  
 DB 32278 AAAATATCTTCTTCAACAGCAATGCGCAAAATTTAGCGAGATGAAAGCTTTAAATG 32337  
 QY 481 TTATATAATATAAGTTCCGGTGGTGAATATATACATATATACATGCAATGCAATAT 540  
 DB 32338 CTTTGCATAAATGCGCTTAACGCCCAAAATATCTCTTTTACCTGCACTAGCAATAC 32397  
 QY 541 CCAAGCCCTTTGAGATGTAACCTTATGCTATTAATGATTTGAAAAACCTTCCT 600  
 DB 32398 CCGGCCCTTTTATGAGATGCAATCTTAAAGCATGCAAGCTTAAAGACGCTT---T 32454  
 QY 601 AAGATATACATAGGCTCTCTGATCATCTTACGAGGGGTTTATGACGATATGCGCGGTG 660  
 DB 32455 AAGCTGATGTGGTATTTCTGATGACACAGGGGTATACACTTTCCCTAGCTGCGGT 32514  
 QY 661 CCTATGGAATTAATCTTTATTTGAAAAACATTTCACTTATGATTAATCTATGTGCGCCA 720  
 DB 32515 GCCTTGGAGACTGTGTGATGATGAAAGCATTTTACTGTACCAAAATATGTCAAGGCT 32574  
 QY 721 GATCATTTGGCTCATATGAAACCTGATGAACTGAAACATTTTGTATGCGGCTCAGGT 780  
 DB 32575 GATCATTTAAACCAAGCTTGAACCTGCAAGAACTTAAATGCTTTGACCCAAATGACAA 32634  
 QY 781 GTTGAATAATCTTTAGGTTCAAAATGTAAGTGGTTACAGCTTCAGAAAGAAATTA 840  
 DB 32635 ATACAAAGGCAATGGTGTATGATGCAATTAATAAGCAATGTAAGCAAAATAATATC 32694  
 QY 841 ATCTAGCAAGAAAGTCTATTTATGCTAAACAGATTAATAAGTGAAGTGTTTTCA 900  
 DB 32695 AACATGTAAGAAAGATTTAGTGGCCAAATAAGACATCAAAAGAGAAATTTTATG 32754  
 QY 901 GAAAAAATATTAACAATAAAGACCTGTAAAGTATGCTATGCTGATGAGGTATTAAT 960  
 DB 32755 GAGGAATCTTACTACCAACGCGCCGCAAAAGTATTAAGTGTATGATATGAGAG 32814  
 QY 961 TTATGGGTAAATTCAGACAGCAAGCTTTATTCAGATGAATTAAT 1007  
 DB 32815 TTTTGGTAAAGTGTACTATAAACTACAAAGAAATAGACTTAAT 32861

RESULT 15  
 AY102621  
 LOCUS 13390 bp DNA linear ECT 08-SEP-2003  
 DEFINITION Campylobacter coli Cj1314c-like protein (Cj1314c) gene, partial cds; Cj1315c pseudogene, complete sequence; Cj1316c pseudogene, complete sequence; and NeuB3 (neuB3), Cj1319-like protein, Cj1320-like protein, Cj1324-like protein, Cj1325-like protein, Pmc (pmc), Pmb (pmb), Pcm (pmb), Pmf (pmf), Pmb (pmb), and Pma (pma) genes, complete cds.  
 VERSION AY102621 AF195053 U25992  
 KEYWORDS  
 SOURCE  
 ORGANISM Campylobacter coli  
 Campylobacter coli  
 Bacteri: Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 Campylobacteraceae; Campylobacter.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 13390)  
 TITLE Query, P., Doig, P., Alm, R.A., Burr, D.H., Kinsella, N. and Trust, T.J.  
 Identification and characterization of genes required for post-translational modification of Campylobacter coli VC167 flagellin  
 JOURNAL Mol. Microbiol. 19 (2), 369-378 (1996)  
 MEDLINE 96423180  
 PUBMED 8825781  
 REFERENCE 2 (bases 1 to 13390)  
 AUTHORS Logan, S.M., Kelly, J.F., Thibault, P., Ewing, C.P. and Query, P.  
 TITLE Structural heterogeneity of carbohydrate modifications affects

JOURNAL serospecificity of Campylobacter flagellins  
 MEDLINE 22294811  
 PUBMED 12406231  
 REFERENCE 3 (bases 1 to 13390)  
 AUTHORS Query, P.M., Doig, P., Alm, R.A., Burr, D.H., Kinsella, N. and Trust, T.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAY-1995) Enteric Diseases, Naval Medical Research Center, 503 Robert Grant Ave, Silver Spring, MD 20910, USA  
 REFERENCE 4 (bases 1 to 13390)  
 AUTHORS Query, P., Ewing, C.P., Moran, A.P. and Trust, T.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-OCT-1995) Enteric Diseases, Naval Medical Research Center, 503 Robert Grant Ave, Silver Spring, MD 20910, USA  
 REFERENCE 5 (bases 1 to 13390)  
 AUTHORS Logan, S.M., Kelly, J.F., Thibault, P., Ewing, C.P. and Query, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2002) Enteric Diseases, Naval Medical Research Center, 503 Robert Grant Ave, Silver Spring, MD 20910, USA  
 REMARK Sequence updated by submitter  
 COMMENT On or Before Jun 2, 2003 this sequence version replaced gi:1049049, gi:11095580.

FEATURES  
 source  
 1..13390  
 /organism="Campylobacter coli"  
 /mol\_type="genomic DNA"  
 /strain="VC167"  
 /db\_xref="taxon:195"  
 /gene="Cj1314c"  
 /complement(1..558)  
 /gene="Cj1314c"  
 /note="probable cyclase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Cj1314c-like protein"  
 /protein\_id="AAW76271.1"  
 /db\_xref="GI:31324510"  
 /translation="MLKRIIPCVLKNLSQVLSIEFQDPTGHTSTMRINARV  
 DELIITDIDSKSGKIDPERSIEDIAKQCMPLTGGITLEDIDKILNIGADKISIN  
 STALEMDPTISKANRSGQCIVCSIDYRKRGQCYVDKRLKRSPLALAEYKK  
 GAGELITVSDDEGRKGYDELAKT"  
 /complement(559..1163)  
 /gene="Cj1315c"  
 /pseudo  
 /complement(1160..2297)  
 /gene="Cj1316c"  
 /note="synonym: psea; involved in synthesis of psea"  
 /pseudo  
 2447..3478  
 /gene="neuB3"  
 2447..3478  
 /gene="neuB3"  
 /note="required for flagella synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /product="NeuB3"  
 /protein\_id="AAW76272.1"  
 /db\_xref="GI:31324511"  
 /translation="MOIGNFNTDKKVFPIIABLSAHTGSLLENALOSIKAAKAGADAI  
 KIOTYTPQSTLNSKEDIKIGLNDKRLVELYSATPYEMHSQIFETTONBGL  
 CRRSPKAEVDKFLKRPDIAYKIASFRANDENFVLIKKKPTIVSGIATBEEL  
 KICEIPKERNKDFLPLKSTVTPYPIAIDEMTKGIVLSKEKNVYVGSDSHSGFLAP  
 VNAVALGARVIEKHMLDLSISEDSKSLDDEFRKAVDAYRQESALGDGKDLDB  
 KALKRVPARSLYAKMDIKKGMFSEENVKSVPSFGIHPKTYQELGKASADIEFG  
 DALKQDDEK"  
 3475..4446  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Cj1319-like protein"  
 /protein\_id="AAW76273.1"



QY 601 AAGATTAACATAGGCTTCTCTGATCTCTAGGGGTTTATGACCTATATGGGGGTG 660  
|||  
Db 8242 AACTTGATGTAGGCTATTCAGATCAACCAAGGATATCATATTTCTTTGACCGCTATC 8301  
|||  
QY 661 CCTTATGAAATTAATTTTATGAAAAATTTCACTTTAGATTAATCTATGTCGGCCCA 720  
|||  
Db 8302 GCCTTAGGGGCAAGCGTATAGAAAAGCATTTTACCTTAGATTAATAATATGTCAGGACCT 8361  
|||  
QY 721 GATCATTGGGCTCAATAGAACTGATGAACCTGAACATCTTTGATTTGGGGTCAAGTGT 780  
|||  
Db 8362 GATCATTAAGGCTCTTTAGAACCTGATGAGCTTCAGAACTTGCACAAAATTTAGAGAA 8421  
|||  
QY 781 GTTGAATAATCTTAGGTCATAATAGTAAAGTGTACAGCTTCAGAAAGGAAAGATATA 840  
|||  
Db 8422 ATCGAAGTGTCTTAGGTGATGGCATATAACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 8481  
|||  
QY 841 ATGTAAGCAAGAAAGTCTATTTATGCTAAACAGAGATTAATAAGGTTGAGTTTTC 900  
|||  
Db 8482 GAAATTCAGAAAGAAAGGCTAGTAGCTTAAGAAATCAAAAAGAGAAATTTTTCAGC 8541  
|||  
QY 901 GAAAAAATTAACAACCAAAAGGCTGTATGATGATCCGATGAGTGTATTAAT 960  
|||  
Db 8542 GAAAGAAAGCTTACACAAAAAGGCTGTATGATGATGATGATGATGATGATGATGATGATG 8601  
|||  
QY 961 TTAATGGGTAAATTTGCAAGCAAGACTTTATTCAGATGAATTAAT 1007  
|||  
Db 8602 TATTTGGGCAAAAAGGCAAGTAAAGATTGAAAGAGATGAGCTTAAT 8648  
|||

Search completed: May 11, 2004, 17:15:55  
Job time : 4316.94 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 456.589 Seconds  
(without alignments)  
9853.151 Million cell updates/sec

Title: US-09-930-440b-7

Perfect score: 1059  
Sequence: 1 atgagatatacatatcgt.....aatgagacaataattatg 1059

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_28Jan04:\*  
1: geneseqn19808:\*  
2: geneseqn19908:\*  
3: geneseqn20008:\*  
4: geneseqn20018:\*  
5: geneseqn20028:\*  
6: geneseqn20038:\*  
7: geneseqn20048:\*  
8: geneseqn20058:\*  
9: geneseqn20068:\*  
10: geneseqn20078:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	100.0	1059	AAA50570	Aa50570 Escherich
2	1059	100.0	1059	AAB28459	Aa28459 Nucleicid
3	1059	100.0	1059	AA153994	Aa153994 DNA encod
4	326.4	30.8	17276	ACA64723	ACA64723 Streptoco
5	324.8	30.7	110000	ABN71527_11	Continuation (12 o
6	321.4	30.3	1023	ABN71527_11	Continuation (12 o
7	317.2	30.0	26281	AA260929	AA260929 Nucleicid
8	163	15.4	110000	AA23109_04	Continuation (5 of
9	107.6	10.2	1050	ABA99765	ABA99765 N. mening
10	107.6	10.2	65632	AA81502	AA81502 N. mening
11	107.6	10.2	110000	AA81490_00	AA81490 N. mening
12	107.6	10.2	349980	AAV21544	AAV21544 Neisseria
13	105.2	9.9	11474	AA53720	AA53720 LPS core
14	105.2	9.9	11474	ABN13665	ABN13665 Campyloba
15	100	9.4	1000	ABK37785	ABK37785 DNA sequ
16	91.4	8.6	999	AA91405	AA91405 N. mening
17	90	8.5	984	ABV72354	ABV72354 Nucleicid
18	70.4	6.6	1123	AA14531	AA14531 H. pylori
19	64.6	6.1	1170	AA14557	AA14557 Human cDN
20	63	5.9	1080	AA37763	AA37763 Human gly
21	63	5.9	1080	AA28458	AA28458 Nucleicid
22	63	5.9	1080	AA15393	AA15393 DNA encod
23	63	5.9	1146	AAA50569	AAA50569 Human sta

24	63	5.9	1230	AA159142	AA159142 Human pol
25	63	5.9	1230	ADB49125	ADB49125 Novel hum
26	63	5.9	1238	AA331141	AA331141 DNA encod
27	63	5.9	1243	AA160928	AA160928 Human pol
28	63	5.9	1268	ABL90630	ABL90630 Human pol
29	61	5.8	1671	AAQ24134	AAQ24134 50 KD sub
30	57.2	5.4	410	AAH35557	AAH35557 Human col
31	52.8	5.0	6394	AAK31374	AAK31374 Signal tr
32	52.8	5.0	6394	AA61278	AA61278 Human gen
33	52.8	5.0	34548	ABL70603	ABL70603 Chemical
34	52.6	5.0	5152	ABL29306	ABL29306 Chemical
35	52.6	5.0	5152	ABL49373	ABL49373 Human pol
36	51	4.8	514	AAH11876	AAH11876 Human cDN
37	50.4	4.8	611	ABX65932	ABX65932 Helicobac
38	50	4.7	547	ABX65406	ABX65406 Helicobac
39	49.6	4.7	6775	ABO67159	ABO67159 Human arg
40	49.4	4.7	6061	ABL32141	ABL32141 Human imm
41	49.4	4.6	3123	ABL28094	ABL28094 Droscophi
42	48.6	4.6	583	ADA72369	ADA72369 Rice gene
43	48.4	4.6	110000	ABA92787_0	ABA92787 Buchnera
44	48	4.5	11650	AA546755	AA546755 Tumour su
45	47.8	4.5	1431	AA237082	AA237082 DNA sequ

## ALIGNMENTS

RESULT 1	AA50570	AA50570 standard; cDNA; 1059 BP.
ID	AA50570	
AC	AA50570;	
DT	19-DEC-2000	(first entry)
DE	Escherichia coli stalic acid synthetase cDNA.	
DE	Escherichia coli stalic acid synthetase cDNA.	
KM	Stalic acid synthetase; neub gene; staliylation; glycoprotein;	
KM	plasmidogen; transletirin; thyrotropin; Na <sup>+</sup> ,K <sup>+</sup> -ATPase; ss.	
OS	Escherichia coli.	
XX	Key	Location/Qualifiers
FT	CDS	1..1041
FT		/*tag= a
XX	W0200052135-A2.	
XX	08-SEP-2000.	
XX	01-MAR-2000; 2000MO-US005313.	
XX	02-MAR-1999; 99US-0122582P.	
XX	08-DEC-1999; 99US-0169624P.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	(UYJO-) UNIV JOHNS HOPKINS.	
XX	(UYWY-) UNIV WYOMING.	
XX	Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA.	
XX	WPI; 2000-572178/53.	
XX	P-PSDB; AAY96102.	
XX	Recombinant production of sialylated glycoproteins using cells in which	
XX	the expression of enzymes, e.g. stalic acid synthetase, involved in the	
XX	sialylation reaction has been altered.	
XX	Disclosure; Page 106-108; 144pp; English.	
XX	The present sequence is that of the Escherichia coli neub gene encoding	
XX	stalic acid synthetase (see AAY96102). The gene was used to identify the	
XX	corresponding human SAS gene (see AAA50569). The invention provides	

SQ Sequence 1059 BP; 396 A; 149 C; 193 G; 321 T; 0 U; 0 Other;

```
Matches 1059; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
```

Db 661 CCTTATGGAATACTTTTATTGAAACAATTCTCACTTAGATAAATCTATGTCGCGCCA 720

Db 1021 TTCAAAATCAGGGGATAATGAGACAAAAATTATTG 10599

ID AAH28459 standard; DNA; 1059 BP.

DT 17-SEP-2001 (first entry)

DE Nucleotide sequence of a bacterial sialic acid synthetase.

KW vaccine; ss.

OS Escherichia coli

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

$$\frac{1}{\sqrt{a}} \frac{d}{dx} \left( \frac{1}{\sqrt{a}} \frac{d}{dx} \right) u = a$$

# IX

PN MO200142492-AL-  
XX

PD 14-JUN-2001

PF 07-DEC-2000; 2000WO-US033136.

PR 09-DEC-1999; 99US-0169839P.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (UTEM) UNIV TEMPLE.

CONFIDENTIAL

PI belindaugu nu, David  
PI vārvīs D;

XX  
DB WPT: 2001-441575/47.

DR P-PSDB; AAB84684.  
YY

PT Cells producing cytidine

PT therapeutics.

PS Example 5; Page 163-1

## CC The specification des

The specification describes a method for manipulating carbohydrate











QY 131 ATAAATTATTTGAGTATTTGACCTTAAGCAGATATCAATATTAACACAGAGAT 190  
 DB 125 AGAACTATTTTCTAAATTTGCTCCCAAGCTGATATCAAAAAGCAATACAGAAACG 184  
 QY 191 TAGATCTCAGTATGAAATGACAAAAGCTGAAATGAGATGACATATCTCCATC 250  
 DB 185 CAGACATCACTGATGATGACGAAAGCTTGAAGTTAAGCTTTGAAAGATCTAGAAA 244  
 QY 251 TAATGGAATTCAGTCAATTAATTTAGATGTTTTTTTCTACCCCTTTGACAGAACT 310  
 DB 245 TGGCTGATATTCGATTTCAAAAGCTGAGACCTTTTCAACCTTTGATTAAGAT 304  
 QY 311 CTATGATTTTTCGATCTTGAACAAAATAAGAAATGCTTCAGGTAGATAT 370  
 DB 305 CATTAAGTTCTTAAATTTCTACAGATATGCCAATTTCAAAATTTCTACAGAAATCA 364  
 QY 371 TGAATTTACCTATCTTGAATAAATAGCCAAAGCTTCGATCCCTGATTAAGAAAATATCA 430  
 DB 365 CTATTTACCTTACCTGAGAAAAGATGGCAGCAAC-----AAAGAAATATTC 415  
 QY 431 TATCAAGAAATGCTACTATGATGATGATTAAGAAAGCTGCTTTCTATTTATTAATA 480  
 DB 416 TTTGAGAGGATATGCGGATATGAGAGATCCATCAAGCGTGAATATTTTACGTCA 475  
 QY 491 ATAAAGTCCGCTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 550  
 DB 476 ATGATCAACCG-----ACATTTCTATTTTACATTTTACAAAGATACCAACCTT 529  
 QY 551 TTGAGATGTAACCTTAATGCTTTATGATTTGAAAAACCTTCCCTAAGATPACA 610  
 DB 530 ACCCTCTCAATTAATTAAGCTTATCAATTTGAAAGATGAAATTTAAAGATTAAACGA 589  
 QY 611 TAGGCTCTCTGATCATTTAGCGGCTTTTATGACGATTTGCGGCGTSCCTTATGAA 670  
 DB 590 TAGGTTATGATCATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 649  
 QY 671 TAACCTTTATGAAAAACATTTCACTTAATTAATTAATTAATTAATTAATTAATTAATTA 730  
 DB 650 CAGAGATATGAAAAACATTTCACTTAATTAATTAATTAATTAATTAATTAATTAATTA 709  
 QY 731 CTCAATTAAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 790  
 DB 710 CCAATGACACCTGATATTTTACGCTTATGATTAAGGCTTCCGATTTGAAACAG 769  
 QY 791 CTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 850  
 DB 770 CTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 829  
 QY 851 GAAAGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 910  
 DB 830 GAAAGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 889  
 QY 911 TAAACAAATAAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 970  
 DB 890 TTAAGGTAAGAGCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 949  
 QY 971 AAATTCAGAGCAAGCTTATTTCCAGATGATTAATTAATTAATTAATTAATTAATTAAT 1030  
 DB 950 AAATTCAGAGCAAGCTTATTTCCAGATGATTAATTAATTAATTAATTAATTAATTAAT 1009  
 QY 1031 A 1031  
 DB 1010 A 1010

RESULT 7  
 AA60929  
 ID AA60929 standard; DNA; 26281 BP.  
 XX  
 AC AA60929;  
 XX  
 DT 30-MAY-2000 (first entry)

XX Nucleotide sequence of a capsular gene cluster of S. suis serotype 2.  
 DB Capsular gene cluster; serotype 2; polysaccharide biosynthesis;  
 XX capsular component; antigen; regulation; chain length determination;  
 KW capsular component; antigen; regulation; chain length determination;  
 KW complement-mediated opsonophagocytosis; serotype-specific detection;  
 KW antigen; vaccine; Streptococcus suis; serotype 2; ORF 2; ORF 2A;  
 KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;  
 KW Cps2L; Cps2M; Cps2N; Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T; ss.  
 XX Streptococcus suis.  
 OS  
 FH Location/Qualifiers  
 Key  
 CDS 2..721  
 FT /tag= a  
 FT /note= "ORF 2Z; encodes AAY68950"  
 CDS complement(822..2079)  
 FT /tag= b  
 FT /note= "ORF 2Y; encodes AAY68951"  
 CDS 2202..2936  
 FT /tag= c  
 FT /transl\_except= (pos: 2916..2918, aa: Xaa)  
 FT /note= "ORF 2X; Xaa is an unspecified amino acid; encodes  
 AAY68952"  
 CDS 3041..4486  
 FT /tag= d  
 FT /note= "Cps2A; encodes AAY68953"  
 CDS 4504..5193  
 FT /tag= e  
 FT /note= "Cps2B; encodes AAY68954"  
 CDS 5203..5880  
 FT /tag= f  
 FT /note= "Cps2C; encodes AAY68955"  
 CDS 5919..6650  
 FT /tag= g  
 FT /note= "Cps2D; encodes AAY68956"  
 CDS 6675..8054  
 FT /tag= h  
 FT /note= "Cps2E; encodes AAY68957"  
 CDS 8089..9258  
 FT /tag= i  
 FT /note= "Cps2F; encodes AAY68958"  
 CDS 9262..10415  
 FT /tag= j  
 FT /transl\_except= (pos: 10057..10059, aa: Xaa)  
 FT /note= "Cps2G; Xaa is an unspecified amino acid; encodes  
 AAY68959"  
 CDS 10808..12176  
 FT /tag= k  
 FT /transl\_except= (pos: 11963..11965, aa: Xaa)  
 FT /note= "Cps2H; Xaa is an unspecified amino acid; encodes  
 AAY68960"  
 CDS 12213..13445  
 FT /tag= l  
 FT /note= "Cps2I; encodes AAY68961"  
 CDS 13583..14581  
 FT /tag= m  
 FT /note= "Cps2J; encodes AAY68962"  
 CDS 14574..15578  
 FT /tag= n  
 FT /note= "Cps2K; encodes AAY68963"  
 CDS 18401..18904  
 FT /tag= o  
 FT /transl\_except= (pos: 18755..18957, aa: Ile)  
 FT /transl\_except= (pos: 18770..18772, aa: Leu)  
 FT /transl\_except= (pos: 18777..18778, aa: Ile)  
 FT /transl\_except= (pos: 18836..18838, aa: Pro)  
 FT /transl\_except= (pos: 18890..18892, aa: Thr)  
 FT /transl\_except= (pos: 18896..18898, aa: Pro)  
 FT /transl\_except= (pos: 18935..18937, aa: Pro)  
 FT /transl\_except= (pos: 18953..18955, aa: Leu)  
 FT /transl\_except= (pos: 18968..18970, aa: Xaa)  
 FT /transl\_except= (pos: 18971..18973, aa: Trp)





```

Db      180 AGAGCAAAATGCTGATGAGGCA-----AACAGCTCATTCAGCAATGCGA 227
Qy      192 AGAATCTCAGTTAGAAATGACAAAAAGCTGAAATGAGTACGATATCTCATCT 251
Db      228 TGTCTCTATTATGAAATTAATGAAACCTTGCCCTCGAATGAAAGATGATTAATT 287
Qy      252 AATGGAATATGACGCTCAGTTAAATTAATGATGTTTTCTTACCCCTTTGAGAGACTC 311
Db      288 AAAAGAAATGCTAGAGCTAAGGATATGTTTTTATGAGTACTCTCTCTGCGAGC 347
Qy      312 TATGATTTTTTATGATCTTTGAAACAAAAATGAAATCCCTCAGTGGATAT 371
Db      348 TGTCTTACGATTTACAGATATGATATTCAGCATATTAATCGGCTTGCGAATGTA 407
Qy      372 GAAATTTACCGTATCTTGAATAAAATGAGCCAGCTCCGATCCCTGATTAAGAAATATCAT 431
Db      408 TAACTACCACTTAATTAATGATGAGCTCTCTT-----TGGTAAGCTATTAATCT 458
Qy      432 ATCAACAGGATGCTACTATGATGAGATTAATAACAGCTCTTCTATTTATTAATAA 491
Db      459 CTCTACCGGCAATGATTTCTATGAAAGCATCAAAAGTGGTAAATTAATTCAGAGAC 518
Qy      492 TAAAGTTCGGTGTGTAATATATACATATTAATGATGCAATCTGATATCCAGCCCTT 551
Db      519 AGGGGTACC-----TTATGCTTTCCTCAGTACGATCAACATTAACCAACCCCTTA 569
Qy      552 TGAGATGTAACCTTAATCTATTAATGATTTGAAAAAAGCTTCCCTTAAGAAATCAT 611
Db      570 CGAAATATGTCATTTGGTGTGATGAAAGATTTATCTAGACCTTTCCAGACGCAATCAT 629
Qy      612 AGGCTTCTCTGATCATCTCAGCGGGTTTATGACAGCTTCGGCGGCTTTATGAAAT 671
Db      630 TGGCTCTGCTGACCTACTTACTATGATGATGCTTCTGAGAGAGATGAGCTTTAGGGCG 689
Qy      672 AACTTTATGAAAAACATTTCACTTATGATTAATCTATGCTGCGCCAGATCATTTGGC 731
Db      690 TTGATTTTAAAGCGCTCACTTACTAGACCGCATGAGTCCGAGTCCGATATGTAAG 749
Qy      732 CTCATTAAGAACCTGATGAACTGAACATCTTTGATTTGGGGCTGAGTGTGAAAAATTC 791
Db      750 CTCTATGATTCGGGATCTTTTAAAGACTCAGAGAGCCCTCATGCTTTAAATTTGGC 809
Qy      752 TTTAGCTTCAATATGTAAGTGTTCAGCTTCAAGAAAGAAAGATTAATGTAAGCAG 851
Db      810 ACGCGCGCGCAAAAAGACACGATATATCGCGGAGAAAGCCAACTAAAGATTTGCGCTT 869
Qy      852 AAGTCTATTATGCTTAAACAGAGATTAATAAAGSTGAGSTTTTTCAGAAAAAATAT 911
Db      870 TGCATCTGCTGAGAGATTAAGACATTAATAAAGAGAACTGTTTGTCCGAGATTACT 929
Qy      912 AACCAAAAAAGACCTGTGTAATG---TATCAGTCCAGTGAAGTGTATTAATTTATTTGG 968
Db      930 ATGGGTAAAGCGCCAGGCAATGAGACTTCAGCTCAACGAATATGAACATTAATTTGG 989
Qy      969 TAAATTTGCAAGCAGACTTATTCAGATCAATTAATTAATTCATCCGATTTCAAAA 1028
Db      990 TAAAGTCTGCTGCTGCAATATTTGCAAGAGTGTCTCAAAATCAAAAACTGATTAATGAATA 1049

```

RESULT 10  
ID AAA81502 standard; DNA; 65632 BP.

XX AAA81502;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_49 SEQ ID NO:49.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;  
XX Meningococcus B; MenB; ds.

```

XX OS Neisseria meningitidis.
XX PN WC20002430-A2.
XX PD 20-APR-2000.
XX PF 08-OCT-1999; 99WC-US023573.
XX PR 09-OCT-1998; 98US-0103794P.
XX PR 30-APR-1999; 99US-0132068P.
XX PA (CHIR) CHIRON CORP.
XX PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Meisigant V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V,
PI Rappuoli R, Pizza M;
XX WPI: 2000-318079/27.
XX DR
XX PT Isolated nucleotide sequences of Neisseria meningitidis which can be used
PT in the diagnosis and treatment of N. meningitidis infection and other
PT Neisserial infections, for example, N.gonorrhoea.
XX PS Claim 7; Page 1331-1350; 1760bp; English.
XX CC The present invention describes methods of obtaining immunogenic proteins
XX from Neisseria genomic sequences. AA81453 to AA82414 represent
XX specifically claimed Neisseria meningitidis genomic DNA sequences,
XX AA81460 to AA81503 and AA825620 to AA825663 represent Neisseria DNA
XX sequences and their corresponding proteins; AA81254 to AA81259 and
XX AA81304 to AA81331 represent PCR primers used in the isolation of
XX Neisseria meningitidis DNA sequences; and AA81332 to AA81452 represent
XX Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
XX used in the exemplification of the present invention. The nucleic acid
XX sequences, protein sequences, and antibodies against them, can be used in
XX the manufacture of a composition. The composition can be used as a
XX medicament (or in the manufacture of a medicament) for treating,
XX preventing or diagnosing infection due to Neisserial bacteria. For
XX example, some of the identified proteins could be components of vaccines
XX against Meningococcus B; against all serotypes; and/or against all
XX pathogenic Neisseriae. Identification of sequences from the bacterium
XX will also facilitate production of biological probes, particularly
XX organism-specific probes. Attempts to make efficacious Meningococcus B
XX vaccines have failed mainly due to antigen tolerance. Multivalent
XX vaccines have also been tried but none have successfully overcome
XX antigenic variability. The provision of further, complete sequences may
XX provide an opportunity to identify secreted or surface exposed proteins
XX that may be presumed targets for the immune system and which are not
XX antigenically variable or at least more conserved than other more
XX variable regions
XX SQ Sequence 65632 BP; 16704 A; 16327 C; 15474 G; 17126 T; 0 U; 1 Other;
XX
XX Query Match 10.2%; Score 107.6; DB 3; Length 65632;
XX Best Local Similarity 47.4%; Prid. No. 6e-13;
XX Matches 483; Conservative 0; Mismatches 504; Indels 33; Gaps 4;
Qy 12 ATATATGTTGCTGTAATAATGTTGCAACATTAATGTTAGTGTATTTTCAGAGAAAT 71
Db 55801 ATTGATTTATCTGTAATAATGCGCATATCATGAAAGGCTCTTTAAAAACACCTTTTGAAT 55860
Qy 72 GATATTAAGCAAAAGAGCGGCTGTATATGAGTAAATTTCCAAACATTTAAAGCTGA 131
Db 55861 GGTGATGCTGCGATATATGAGCGCTGAAGTGTATTAACATCAACACACATGTTGA 55920
Qy 132 TAAATTAATTTACGATATTCACCTAAGCAGATATCAATTAATAAACAAGAGATT 191
Db 55921 AGAGAAATGCTGATGAGCCCA-----AACAGTCAATTCAGCAATGAGCA 55968
Qy 192 AGAATCTCAGTTAGAAATGACAAAAAGCTGAAATGAGTACGATATCTCATCT 251
Db 55969 TGTCTCTATTATGAAATTAATGAAACCTTGCCCTCGAATGAAAGATGATTAATT 56028

```

```

QY 252 AATGATATGAGTCACTTTAAATTAGATGTTTCTACCCCTTTGACGAGATC 311
DB 56029 AAAAGATACGTAGAGTAAAGGATATGATTTTATCAGTACTCTCTCTCTGTCAGC 56088
QY 312 TATGATTTTAAAGATCTTTGAAACAAAATAATGAAAAATCCCTTCAGTGAATTT 371
DB 56089 TGCTTACGATTAACAGTATGATATTTCCAGCATTAATAATGCGCTGCGCAATGTA 56148
QY 372 GAATTTACCGATCTGAAAAAATAGCCAGCTCCGATCCCGATTAAGAAATTAATCAT 431
DB 56149 TAACCTACCTTAATTAATTAATCTGCTGCTCTCTT-----TGTAAAGCTTATATCT 56199
QY 432 ATCAACAGATGAGTCTACTATGATGATATAAAGCTGCTTCTATTTTAAATAA 491
DB 56200 CTCTACCGGATGATATCTATGAAAGCATCAAAAATCGGTGAATATTTGGAAGAC 56259
QY 492 TAAAGTCCCGTGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 551
DB 56260 AAGGGTAC-----TTATGCTTGTCTCACTGACCAATCAATCAACCCCTTA 56310
QY 552 TGAAGATGTAACCTTATGATTAATGATTAATGAAAAAATCCCTTCAGTGAATCAT 611
DB 56311 CGAAGATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56370
QY 612 AGGCTTCTGATCATTTAGCGGTTTAAAGAGCTATTGCGCGGCTGCTTATGGAAT 671
DB 56371 TGCGCTTCTGACCAATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 56430
QY 672 ACCTTATGAAAAAATCTTCACTTAATTAATTAATTAATTAATTAATTAATTAAT 731
DB 56431 TTGATTTAGAGTCTTACTTACTGACCGCATGATGATGATGATGATGATGATGAT 56490
QY 732 CTCAATGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
DB 56491 CTCTAGATCCGATCTTCTTAAAGAGCTCAAGCAAGGCTCATGCTTAAATTTGC 56550
QY 792 TTTAGTCAAAATGTAAGTGTACAGCTTCAGAAAGAAATTAATTAATTAATTAAT 851
DB 56551 ACCGCGGCGGCAAAAGACAGATATCGCGGAGAAAGCAATTAATTAATTAATTAAT 56610
QY 852 AAAGTATATATGATTAACACAGATTAATAAAGGATGATGATGATGATGATGATGAT 911
DB 56611 TGCATCTGCTGACAGATTAAGACATTAATAAAGGATGATGATGATGATGATGAT 56670
QY 912 AACAAACAAAGACCTGCTATG--TATGATGCTGATGATGATGATGATGATGATG 968
DB 56671 ATGGGTAAAGCCCAAGGCAATGAGACTTCAGCTCAACGAATATGAACATTAATTTG 56730
QY 969 TAAATGACAGACGATTTATTCAGATGAATTAATTAATTAATTAATTAATTAATTA 1028
DB 56731 TAAAGTCTGCTGCTGCAATTTGCAAAAGTCTGCAATTAATAAAGTATTAATTA 56790

```

## RESULT 11

AAA81490\_00/c  
 WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aa81490

Fragment Name	Begin	End
AAA81490_00	1	110000
AAA81490_01	100001	210000
AAA81490_02	200001	310000
AAA81490_03	300001	410000
AAA81490_04	400001	510000
AAA81490_05	500001	610000
AAA81490_06	600001	710000
AAA81490_07	700001	810000
AAA81490_08	800001	910000
AAA81490_09	900001	1010000
AAA81490_10	1000001	1110000
AAA81490_11	1100001	1210000
AAA81490_12	1200001	1310000
AAA81490_13	1300001	1410000
AAA81490_14	1400001	1437668

```

ID AAA81490 standard; DNA; 1437668 BP.
XX
XX AAA81490;
AC
XX
XX 04-DEC-2000 (first entry)
DT
XX
XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen, vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; MenB; de.
XX
XX Neisseria meningitidis.
OS
XX
XX MO200022430-A2.
XX
XX 20-APR-2000.
PD
XX
XX 08-OCT-1999; 99MO-US023573.
PF
XX
XX 09-OCT-1998; 98US-0103794P.
PR
XX 30-APR-1999; 99US-0132068P.
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Mastiani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappoli R, Pizzi M;
PI
XX
XX MPI; 2000-318079/27.
DR
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be used
PT in the diagnosis and treatment of N. meningitidis infection and other
PT Neisseria infections, for example, N.gonorrhoea.
PT
XX
XX Claim 7: Page 866-1272; 1760pp; English.
XX
XX
XX The present invention describes methods of obtaining immunogenic proteins
CC from Neisseria genomic sequences. AAA81453 to AAA82414 represent
CC specifically claimed Neisseria meningitidis genomic DNA sequences;
CC AAA81260 to AAA81303 and AAA825620 to AAA825663 represent Neisseria DNA
CC sequences and their corresponding proteins; AAA81254 to AAA81259 and
CC AAA81304 to AAA81321 represent PCR primers used in the isolation of
CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent
CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
CC used in the exemplification of the present invention. The nucleic acid
CC sequences, protein sequences, and antibodies against them, can be used in
CC the manufacture of a composition. The composition can be used as a
CC medicament (or in the manufacture of a medicament) for treating,
CC preventing or diagnosing infection due to Neisseria bacteria. For
CC example, some of the identified proteins could be components of vaccines
CC against Meningococcus B; against all serotypes; and/or against all
CC pathogenic Neisseriae. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious Meningococcus B
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions
CC
XX
XX Sequence 1437668 BP; 344338A; 353206C; 385074G; 355045T; 0U; 50ther;
SQ

```

Query Match 10.2%; Score 107.6; DB 3; Length 110000;

Best Local Similarity 47.4%; Pred. No. 6.5e-13;

Matches 483; Conservative 0; Mismatches 504; Indels 33; Gaps 4;

```

QY 12 ATATATCGTGTGTAATTTGTTGCAACCATATATGATGTTGATATGCAAGAAAT 71
DB 78357 ATGATATCTGTGAATTCGAGATCATATCAAGAGCTTTAAATAACGCTTTGAAAT 78298

```







```

Db      78189 TGCTCTATTATTAGTAATTTATGGAACCTGGCCCTGAAATGAAAGATAGATTAAT 78130
Qy      252 AATGGAATATGACGATCGATTTAAATTTAGATTTTCTACCCCTTTTGAAGAGATC 311
Db      78129 AAAAGATACGATGAGAGTAAAGGTATGATTTTATGAGTACTCTTCCTGTCAGC 78070
Qy      312 TATGATTTTTCGATCTTTGAAACAAAATAATGAAAAATCCCTCAGGTGATAT 371
Db      78069 TCGTTTCGATTAACGATATGATATCCAGCATATAAATGCGCTCTGCGAATTA 78010
Qy      372 GAATTTACCGATCTTGAATAAATAGCCAGCTCCGATAGAAAATATATCAT 431
Db      78009 TAATACCCATTAATTAACGTGGCTCTT-----TGGTAAGCTATATTC 77959
Qy      432 ATCAACGGAATGGCTCTATGATGATTAACATGCTTTCTATTTTATTAATA 491
Db      77958 CTCTACCGGATGAAATCTATGAAAGCATAAAAGTCGGTAGAAATATTCGAGAGC 77899
Qy      492 TAAAGTCCGGTTGGTAAATATTAACATATACATTCGAAATATCCAGCCCT 551
Db      77898 AGGGGTACC-----TATGCTTGGCTTCACTGATACCAACATTAACCAACCCCTTA 77848
Qy      552 TGAGATGTAACCTTAATGCTATTAATGATTTGAAAAACACTCCCTTAAGATACAT 611
Db      77847 CGAAGATGTTGATGGGTGGTAAAGCATTTATCTGAACCTTTCAGAGCAATCAT 77788
Qy      612 AGGCTCTCTGATCACTTACGCGGTTTATGACGCTATTCGCGGCTTATGGAAT 671
Db      77787 TGGCTGTCTGACCACTTACATTAATGATGCTTGTGAGGACATAGCTTTAGCGG 77728
Qy      672 AACTTTATGAAAAAATTCCTTCACTTAAGATTAATCTATGCTGCGCCCATTTGCG 7731
Db      77727 TTCCATTTTGAAGAGCTCACTTACTGACCGCATGATCGCCAGCTCCGATATTGATG 77668
Qy      732 CTCATTAAGACCTGATGACGAAACATCTTGTATGAGGGTTCAGGTGTGTAATAATC 791
Db      77667 CTCATTAATCGGATCTTTTAAAGACTTAAAGCAAGCAAGGCGCTCATGCTTTAAATGGC 77608
Qy      792 TTTAGTCAATAGTAAGGTGTTACGCTTCAGAAAGAAATTAATCTAGCAAG 851
Db      77607 ACGGCGCGGCAAAAAGCAAGATTAATCGCGGAGAAAGCACTAAAGATTGCGCTT 77548
Qy      852 AAAGCTATATAGCTTAAACAGAGATTAATAAAGGTGAGGTTTTCAGAAAAAATAT 911
Db      77547 TGCATCTGTCTGACGATTAAGCACTTAAAGAGAACTGTGTCCGAGATTAACCT 77488
Qy      912 AACACAAAAGACCTGGTATG---TATCATGCCATGAGTGTATTAATTTATGGG 968
Db      77487 ATGGGTTAAAGCCCGAGCATGAGACTTCAAGCTCAAGAAATATGAACATTATTTGG 77428
Qy      969 TAAATTTGACAGACAGACCTTATTCAGATGAATTAATTAATCATAGCGAATCAAAA 1028
Db      77427 TAAAGTGTGCTGCTTGCATATATTCGAAAGGTGCTCAATAACAAAACGTATATGAATA 77368

```

## RESULT 13

AAAS3720 standard; DNA: 11474 BP.

AAAS3720;  
 AC AAA53720;  
 XX 15-SEP-2003 (revised)  
 DT 22-DEC-2000 (first entry)  
 XX

LPS core biosynthesis locus.

XX Biosynthetic locus; biosynthesis; lipid A biosynthesis;  
 KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;  
 KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;  
 KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;  
 KW immunity; immunogen; ganglioside; ds.  
 XX

```

OS Campylobacter jejuni, O4384.
FH Key Location/Qualifiers
FT CDS 350..1237
FT /tag= a
FT /product= "acetyltransferase"
FT /note= "Open reading frame 2b"
FT CDS 1234..2490
FT /tag= b
FT /product= "glycosyl transferase"
FT /note= "Open reading frame 3a"
FT CDS 2786..3955
FT /tag= c
FT /product= "glycosyl transferase"
FT /note= "Open reading frame 4b"
FT CDS 4025..5068
FT /tag= d
FT /product= "beta 1,4-GalNAc transferase"
FT /note= "Open reading frame 5a"
FT CDS 5054..5959
FT /tag= e
FT /product= "beta 1,2- galactosyltransferase"
FT /note= "Open reading frame 6a"
FT CDS 6048..6823
FT /tag= f
FT /product= "CstII sialyltransferase"
FT /note= "Open reading frame 7a"
FT CDS 6924..7964
FT /tag= g
FT /product= "sialic acid synthase"
FT /note= "Open reading frame 8a"
FT CDS 8020..9079
FT /tag= h
FT /product= "sialic acid biosynthetic enzyme"
FT /note= "Open reading frame 9a"
FT CDS 9076..9741
FT /tag= i
FT /product= "CMP-sialic acid synthetase"
FT /note= "Open reading frame 10a"
FT CDS complement (10554..11366)
FT /tag= j
FT /product= "lipooligosaccharide biosynthetic enzyme"
FT /note= "Open reading frame 12a"
XX W0200046379-91.
XX 10-AUG-2000.
XX 01-FEB-2000; 2000MO-CA000086.
XX PF 01-FEB-1999; 98US-0118213P.
XX PR 31-JUN-2000; 2000US-00495406.
XX XX (CANA ) NAT RES COUNCIL CANADA.
XX PA Gilbert M, Wakarchuk WM;
XX PI WPI: 2000-524418/47.
XX DR P-PSDB; AA97200, AA97201, AA97202, AA97203, AA97204, AA97205,
XX DR AA97206.
XX PT Novel glycosyltransferase polypeptides and polynucleotides useful for
XX PT biosynthesis of ganglioside and ganglioside mimics, as diagnostic
XX PT reagents and as immunogen for producing antibodies.
XX PS Claim 1; Page 86-90; 120pp; English.
XX A reaction mixture for the synthesis of a sialylated oligosaccharide is
XX CC useful for synthesizing sialylated oligosaccharides such as ganglioside,
XX CC lysoganglioside or their mimics. Glycosyltransferases are useful for
XX CC chemo-enzymatic synthesis of oligosaccharides, including gangliosides and
XX CC other oligosaccharides that have biological activity. The enzymes and
XX CC nucleic acids that encode them are useful for studies of the pathogenesis

```





Thu May 13 11:53:11 2004

us-09-930-440b-7.ing

Page 16

```
QY 256 GAATATGAGTACGATTAAATTAGATGTTTTTCTACCCCTTTGAGAGACTGATT 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 GAATACGTAGAGAGTAAAGGATGATTTTATACAGTACTCCTTCTCTGCGACGCT 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 GATTTTATAGCATCTTTGAAACAAAATATGAAATCCCTTCAGTGAGTTATGAT 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 TTACATTACAAAGTATGATTTTCCACATATATAATCGGCTCGGCAATGTATAC 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 TTACCGTATCTTGAATAATACCAAGCTTCGATCCCTGATAGAAATATCATCA 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 TACCATTAATTAACCTGGTGGCTCTTT-----TGGTAAAGCTTATATCTCT 329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 436 ACAGAAATGGCTACTATTGATAGATAAACAGCTGTTTCTATTTTATATAATAA 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 ACCGGCAATGAAATCTATTGAAAGCAATCAAAAAGTCGTAGAAATTTTCGAGACAGG 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 496 GTTCGGTGTGTAAATATTAACATATATACATGCAATGCAATATCCAGCCCTTAG 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 GTACCTTATGCTTGTCT-----CACTACCAACATCTACCCCAACCCCTTACGA 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 556 GATGTAAACCTTAATGCTATTATATGATTTGAAAAACCTTCCCTAAGATTAACAGG 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 441 GATGTGAGTGGGTGATGAGCAATTTATCTGAAGCCTTTCAGACGCAATCATGSC 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 616 TTCTCTGATCATTCTAGCGGGTTTATGCAAGCTATGCGGCGGTGCTTANGAATACT 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 CTGCTGACCATACCTTATAGTACTGATGCTTGTAGAGCAGTAGCTTAGCGGTG 560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 676 TTTATGAAAAACATTTCACTTATAGTAAATCTATGCTGCGCCAGATCATTTGGCCTCA 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 561 ATTTTAGAGCGTCACTTACTAGACGCAATGATCGCCGAGTGGATATGTATGCTCT 620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 736 ATAGAACCTGATGAGTGAACATCTTTGATGCGGTGAGGTGTTGAAAAATCTTTA 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 621 ATGATCCGGATACCTTTTAAAGCTCAGCAAGGCGCTCATGCTTAAAAATTGGCAGCG 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 796 GGTTCAAATAGTAAAGTGTACAGCTTACAGAAAGAGATTAATCGTACAGAGAAAG 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 681 GCGCGCAAAAAAGACAGATTAATCGCGGAGAAAGCCAACTAAAGATTTGCGCTTGCA 740
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 856 TCTATATAGCTTAAACAGAGATTAATAAAGGTGAGGTTTTTTCAGAAAAAATATACA 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 741 TCTGTCTAGCAGATTAAGCACTTAAATAAAGAGAACTGTTCGCGAGATAACCTATGG 800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 916 ACAAAAAGACCTGTATATG---TATCAGTCCGATGAGGTGTATATTTATTTGGTTAA 972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 801 GTTAAACGCCCAAGGCAATGAGACTTCAGCGTCAACGAATATGAAACATTTTGGTTAG 860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 973 ATTGCAGAGCAACTTATTCAGATGATTAATTAATTCATAGCGAATTCAAAAA 1028
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 861 GTGCTGTGCTGCAATATTCGAAAGGTGCTCAATCAAAAAAATGATATGAAAT 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 10, 2004, 08:08:15  
Job time : 464.589 secs

Thu May 13 11:53:12 2004

us-09-930-440b-7.rst

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 : Search time 3062.04 Seconds  
(Without alignments)  
10327.779 Million cell updates/sec

Title: US-09-930-440b-7  
Perfect score: 1059  
Sequence: 1 atgaggaacataatataatgc.....aattgagacacaaatattatg 1059

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gsa\_hum:\*  
18: em\_gsa\_hiv:\*  
19: em\_gsa\_pln:\*  
20: em\_gsa\_vic:\*  
21: em\_gsa\_fun:\*  
22: em\_gsa\_mam:\*  
23: em\_gsa\_mus:\*  
24: em\_gsa\_pro:\*  
25: em\_gsa\_rtd:\*  
26: em\_gsa\_phg:\*  
27: em\_gsa\_vir:\*  
28: gb\_gsa1:\*  
29: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	8.3	649	14	CF998654 ACBNCOURT
2	79.4	7.5	1200	13	EX437758 EX437758
3	78.8	7.4	1435	11	BC050955 Danto rer
4	73.8	7.0	876	14	CD327087 ACBNCOURT

5	71.8	6.8	508	14	CA550431
6	69.4	6.6	545	12	B1377301
7	68.6	6.5	715	12	B185296
8	68.2	6.5	868	14	CD301354
9	68.2	6.4	676	12	BM095216
10	67.6	6.3	673	14	CF360452
11	67.6	6.3	1225	29	CNS0161D
12	66.4	6.3	613	13	BQ600407
13	66.2	6.3	643	12	B069790
14	66.2	6.3	678	14	CE530032
15	65.6	6.2	469	14	CB733943
16	65.6	6.2	1070	9	AL544111
17	65.2	6.2	417	10	AM416754
18	65.2	6.2	886	28	BH177277
19	65.2	6.2	886	29	CNS07JDX
20	65.2	6.2	1101	29	CNS0039G
21	65	6.1	1201	13	EX335216
22	64.8	6.1	574	14	CA417649
23	64.6	6.1	590	12	BQ014133
24	64.6	6.1	710	10	BR448706
25	64.6	6.1	777	12	BG568099
26	64.2	6.1	576	14	CD365322
27	63.6	6.0	579	14	CA314981
28	63.4	6.0	653	14	CB427711
29	63	5.9	551	12	BH179516
30	63	5.9	569	10	AM249532
31	63	5.9	576	12	BM792591
32	63	5.9	577	13	BU626163
33	63	5.9	577	13	CA429806
34	63	5.9	579	13	BQ773353
35	63	5.9	604	14	CA434954
36	63	5.9	608	14	CD671274
37	63	5.9	610	12	BQ001230
38	63	5.9	613	13	BQ676733
39	63	5.9	677	14	CA411576
40	63	5.9	677	13	BU687286
41	63	5.9	689	12	BM976474
42	63	5.9	694	13	BQ446501
43	63	5.9	698	12	BM982564
44	63	5.9	769	13	BU618375
45	63	5.9	779	14	CA417480

## ALIGNMENTS

CA550431 C0830806-  
B1377301 BPLG3\_001  
B185296 UNL-P-FN-  
CD301354 ACBNCOURT  
BM095216 fvg1g11.x  
CF360452 821784 MA  
AL106171 Drosophila  
BQ600407 MI-P-E7-A  
B069790 B1069790  
CE530032 UI-H-FR2-  
CB733943 AMGNIC:M  
AL544111 AL544111  
AM416754 52368 MAR  
BH177277 008\_L\_22-  
AL614235 T3 end of  
AL063921 Drosophila  
BX335216 BX335216  
CA417649 UI-H-FE0-  
BQ014133 UI-H-ED1-  
BR448706 7n93d07.x  
BG568099 602586980  
CD365322 UI-H-FR2-  
CA314981 UI-H-FE0-  
CB427711 603306 MA  
BG179516 602328112  
AM249532 2821389.3  
BM792591 K-EST0072  
BU626163 UI-H-FG1-  
CA429806 UI-H-FH1-  
BQ773353 UI-H-FE0-  
CA434954 UI-H-DH0-  
CD671274 fg02906.x  
BQ001230 UI-H-DH1-  
BU676733 UI-H-ED1-  
CA411576 UI-H-ED1-  
BU687286 UI-H-ED1-  
BM976474 UI-H-ED1-  
BQ446501 UI-H-ED1-  
BM982564 UI-H-ED1-  
BU618375 UI-H-FH1-  
CA417480 UI-H-FE0-

RESULT 1  
LOCUS 649 bp mRNA linear EST 25-NOV-2003  
DEFINITION AGENCOURT\_16388159 NIH\_ZGC\_7 Danto rerio cDNA clone IMAGE:7039969  
5' UTR RNA Sequence.  
ACCESSION CF998654  
VERSION CF998654.1 GI:38519505  
KEYWORDS EST.  
SOURCE Danto rerio (zebrafish)  
ORGANISM Danto rerio  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Cypriniformes; Cyprinidae; Danto.  
REFERENCE  
1 (bases 1 to 649)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
NHLBI National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Len Zou, Harvard  
cDNA Library Preparation: Open Biosystems  
cDNA Library Arrayed by: The J.M.A.G.B. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be





```

/dev_stage="adult"
/lab_host="DH10B_(phage-resistant)"
/clone_lib="NICHD_XGC_Eyal"
/name="Organ:EYE; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Primer: Oligo dT
Average insert size 2.3 kb. Constructed by ligation
Technologies. Note: This is a Xenopus Gene Collection
(XGC) library."

```

ORIGIN

	Query Match	Similarity	48.5%	Score	73.8	DB 14	Length	876
	Best Local	Similarity	48.5%	Pred. No.	0.00012			
	Matches	204	Conservative	0	Mismatches	217	Indels	0
QY	508	AATATTACAAATTATACATTGCGAATCTGAAATATCCAAAGCCCTTTGAGATGTAATACCTT						567
Db	39	AACCTTTGGCTTTCTACAGTGCACAAAGTGCATACCCCACTTTACCAAGGAGCTCAACCTTC						98
QY	568	AATGCTATTATGATGATTGTAAGAAAACATCTCCCTAAGATACATAGAGCTCTCTGATCAT						627
Db	95	TGATTCATTAAGAGAGTTTCACTGATCTTCTCCAGATATCCAAATGGCTACTCTGGATCAT						158
QY	628	TCATAGCGGGTTTATGCACTATTGCGGGGTGCTTATGGAATACCTTTATTGTAAGAAA						687
Db	159	GAACCTGGGATCTGCTATCTCACTGTCAGACTGTGGTATGGAGGCCAAGGTGGAGGCCG						218
QY	688	CATTTCACCTTATGATTAATATCTATGCTGAGCCCAAGCTATTGGGCGCTCAATAGAACTGAT						747
Db	215	CATGTATACACTGACCAAACTTGAAAGGCAATGATATCACAGAGTTGCGTGGAGCCAAAGT						278
QY	748	GAACCTAAACATCTTTGTAATGAGGAGTGTGTAAGAAATCTTTAGCTCAATATAGT						807
Db	279	GAATTCACAGAAATTATGTCAGCTTCATCTCAGACTTGTAGAGAAAGAAAGGAGCTCCACTGTC						338
QY	808	AAATGTGGTTACAGCTTCAGAAAGGAAAGAAATTAATTCGTAGCAAGAAAGCTATATAGCT						867
Db	339	AAATAGCTGTGGCATGTGATATTTGGCCCTGTCAACAAAGCTCGGTATATCAAGTGTGGCC						398
QY	868	AAATACAGAGATTAAGAAAAGGTGAGGTTTTTTCAGAAAAAATATTAACAACAAAAGACT						927
Db	399	AAATGTATAGATTCACAGCTGGCACAGGTGTCAGCTCAATATGCTTAACGGTGAAGATGGCT						458
QY	928	G 928						
Db	459	G 459						

RESULT 5			
CA550431			
LOCUS	508 bp	mRNA	linear
DEFINITION	C0830306-5N N1A Mouse BlascoGy8t cDNA Library (Long) Mus musculus		
	cDNA clone N1A:C0830E06 IMAGE:30027893 5', mRNA sequence.		

ACCESSION	CA550431	GI:25094538
VERSION	CA550431.1	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	

TITLE	REFERENCE
JOURNAL	AUTHORS
COMMENT	
Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)	Ko, M.S.H.
Unpublished (2001)	Piao, Y., Kargali, G.C., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and Euhairiot, Euhairiot; Chandra; Chandra; Versteeg; Versteeg; Mammalia; Euhairiot; Rodentia; Scleroglossa; Muridae; Murinae; Mus. 1 (bases 1 to 598)
Other_ESTS: C0830E06-3	

Contact: Dawood S. Dunderina  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6620, USA  
Email: cdna@sgsun.gic.nia.nih.gov  
Plate: C0830 row: E column: 06  
Seq primer: M13 Reverse  
High quality sequence stop: 508

POLYA=No.	Location/Qualifiers
FEATURES	1. 508
source	/location="Muc

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taabEST:C0830B06-5N"
/db_xref="tacon:10090"
/clone="NTA:C0830B06 IMAGE:30027893"
/tissue_type="Blastocyst"

```

```

/dev stage="3.5-gpc"
/lab host="DH108"
/cclone lib="NIA Mouse Blastocyst cDNA Library (long)"
/ncore-vecloc:rsPORT1 (Invitrogen) ; Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.gsc.nia.nih.gov/cDNA/). This is
a long-term transcript-enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11541493]). Total RNAs were
extracted from a pool of 20 blastocysts. Double-stranded
cDNAs were synthesized with an Oligo(dT) primer

```

**Library:** 5'-GACATGCTTCATGATCCGACGACGCGCCGCCCTCTTTTCTTTT-3' from 0.2 µg of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to *l*-one-linker *l*-Sal-I, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer Sal4-5. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes, and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Julian Piao (MIA).

ORIGIN

Query Match	6.8%	Score 71.8;	DB 14;	Length 508;
Best Local Similarity	50.9%;	Pred. No. 0.0034;		
Matches 200;	Conservative 0;	Mismatches 187;	Indels 6;	Gaps 1

```

QY      12 ATATATCGTTCGTGAATATGGTGTGAACCATATATGTGATGTGATATTTGCAAGGAAT 71
Db      40 ATTATATATTCGAGAAATGAGGTGGCAATCATATATCATCTATAGATTAAGCATTTAAAT 99
QY      72 GATATTTAAAGCCAAAGGCGCGGTGTATAGCGGTAAATTTCCAAATTTAAAGCTCA 131
Db      100 TATTGAAACTGCTGCAGAAAGTGTGGGCCCATGCACTTAACTACAAACATATACAGCCGA 159
QY      132 TAAA-----TTAATTTCCGCTATTTGCACCTTAAGCAGAGATCAATATAAAAACAGG 185
Db      160 TACATACACCTTAATATCATAGAGGTGCACCTTTGCATATTCAGATGAAAGCTCTTGTG 219
QY      186 AGAATTAGATCTCAGTTAGAAATGACAAAAAGCTTGAAATGAAGTATGACATTATCT 245
Db      220 GAAAAACAGGAATTTGTATGAACATATATCAAGAACATATATCTCATGGAAATGGCATGA 279
QY      246 CCATCTATAGGAATATGCAAGTCAGTTTAAATTAGATGTTTTCTTACCCCTTTTGACGA 305
Db      280 AGCTGTTATTTAAATGTGACGCTGATACAAATATCATTTATTTTCTACACCTTTTGATGA 339
QY      306 AGACTCTATGATTTTATTTAGCATCTTTGAAACAAAAAATATGAAAAATTCCTTCAGSTGA 365
Db      340 CAGTGCAGGTGATTTTTTTTGAAGAAAGTTAAATTCGCCAGCCTAATAAAATCGCGTCTTTTGA 399
QY      366 GTTATGTAATTCAGCTATCTTGAAAAAAAATAGC 398
Db      400 AAATPACTATTTTACAGTTATTTAAAAAAGGTAC 432

```

RESULT 6  
BI377301



**ପଞ୍ଚମ**

B1377301 545 bp DNA library linear EST 26-KX-2003  
 BFLG3 001104 Amphioxus 5-6 cDNAs from the same convention: BFLG  
 or MPMG4968 Branchiostoma floridae cDNA clone MPMG4981399 5',  
 mRNA sequence.  
 B1377301  
 B1377301.1 GI:30912237  
 EST.  
 Branchiostoma floridae (Florida lancelet)  
 Branchiostoma floridae  
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomida;  
 Branchiostoma.  
 1 (bases 1 to 545)  
 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,  
 Herwig, R., Vingron, M. and Lehrach, H.  
 New evidence for genome-wide duplications at the origin of  
 vertebrates using an amphioxus gene set and completed animal  
 genomes  
 Genome Res. 13 (6A), 1056-1066 (2003)  
 22683279  
 12795346  
 Contact: Panopoulou G  
 Laboratory 145, dept Lehrach  
 Max-Planck-Institut fuer Molekulare Genetik  
 Ihnestr. 63-73, D-14195 Berlin, Germany  
 Tel.: +49 30 8413 1235  
 Fax: +49 30 8413 1128  
 Email: panopoulou@molgen.mpg.de  
 The library was characterised by oligonucleotide fingerprinting  
 (ONFP) to reduce sequencing redundancy. According to the ONFP  
 procedure, clones giving the same hybridisation pattern with a  
 battery of 200 8mer oligonucleotides are grouped into clusters. One  
 clone per cluster is selected for sequencing. The size of each  
 cluster is an indicator of the frequency of a transcript in the  
 analysed library. The cluster size as well the coordinates of the  
 rest of the clones assigned to the same fingerprint cluster as the  
 clone from which the above EST is generated is available at  
 the amplicon project site at: <http://www.molgen.mpg.de/amphioxus/>  
 Clones and filters are distributed via the Resource Center/Primary  
 Database of the German Genome Project (<http://www.izpdp.de>)  
 PCR Primers  
 FORWARD: 5' CCCGAGCTTACACCTTAACTCCGGCTG 3' (M13SP)  
 BACKWARD: 5' GCGATTACGGACACTGATGAGGGGATGG 3' (M13FSP)  
 Insert length: 1200 Std Error: 0.00  
 Seq primer: 5'-CCGCTCCGAGATTCGGGCT-3' pSPORT3/86  
 High quality sequence. Stop: 545.

Query Match	6.6%	Score 69.4	DB 12	Length 545
Best Local Similarity	48.2%	Pred. No. 0.0089		
Matches 196	Conservative	0	Mismatches 211	Indels 0
			Gaps	0

465 ACAGTCTGTTTCTATTTTTATAAATAATPAAGTTCGGTTGGTATATTACATATTACA 524

RESULT 7					
B1185296/c					
LOCUS	B1185296	715 bp	mRNA	linear	EST 10-JUL-2001
DEFINITION	UNI-P-FN-cu-a-06-0-0-UNL.s1	UNI-P-FN	scrofa	CDNA	clone
	UNI-P-FN-cu-a-06-0-0-UNL.3'				mRNA sequence.

REFERENCE	1 (bases 1 to 715)
AUTHORS	Caetano, A. R., Johnson, R. K. and Pompl, D.
TITLE	Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
JOURNAL	Mamm. Genome 14 (1), 65-70 (2003)
MEDLINE	22419904
PMID	1253269
COMMENT	Contact: Pompl, D.

FEATURES	Location/Qualifiers
source	1. .715

/notes: Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI. The UML-P-PN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase.

of the pig estrous cycle. This library was derived from the library UNL-P-2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Ronaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

TAG\_TISSUE=porcine ovarian follicles  
TAG\_LIB=UNL-P-FN  
TAG\_SEQ=CACACT

## ORIGIN

Query Match 6.5%; Score 68.6; DB 12; Length 715;  
Best Local Similarity 48.7%; Pred. No. 0.011;  
Matches 185; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

521 TACATGCAATAGTGAATATCCACGCCCTTGGAGATGAACCTTAATGCTATTATG 580  
572 TCCAGTGCACCGGCGGTACCCACTCCACCTGAGACGTCACCTGGCGTATCTCGG 513  
581 ATTGAAAAACCTTCCCTAGATTAACATAGGCTTCTGATCTCTAGCGGCTTT 640  
512 AATATCAGAAAGCTTCTCCGATTCATCCATAGGATATCTGGGCAATGAACGAGATG 453  
641 ATGACGCTATGTGGCGGCTGCTTATGATATCTTATGAAAAACATTCACTTATG 700  
452 CAATATCTGTGCTGCGCTGCTGCTGAGACCAAGCTCTGAGCTCACTAATCTTGG 393  
701 ATAAATCTATGTGTGCGCCAGATCATTTGGCTCAATAGAACCTGATGAACCTGAAC 760  
392 ACAAGCTTGGAGGAGGAGTACACTCGGCTCGTGAAGCTGAGAGCTGGCCGAC 333  
761 TTGTATGTGGGCTGAGGTGTGTAATAATCTTAGATCAAAAGTAAGTGTATGAC 820  
332 TGTGTGGGTCTGTGCGCTCTGTGAGAGGCTTGGGCTCCCAACCAAGCACTGTGCG 273  
821 CTTCAGAAAGAGAAATAAATGCTAGCAAGAAAGTCTATATAGCTAAACAGAGATTA 880  
272 CCTGTGAGATGCGCTCAACGAGAAAGCTGGCAAGTGTGTGCGCAAGTGAATTC 213  
881 AAAAGTGAAGTTTTTCA 900  
212 CGAAGGCACTGTCTTAACA 193

RESULT 8 868 bp mRNA linear EST 27-MAY-2003  
CD301354  
LOCUS AGENCODE 14234689 NICHD\_XGC\_Eyel Xenopus laevis cDNA clone  
DEFINITION IMAGE:6959626 5', mRNA sequence.  
ACCESSION CD301354  
VERSION CD301354.1 GI:31081149  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodidae; Xenopus.  
1 (bases 1 to 868)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabs-remail.nih.gov  
Tissue Procurement:  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov

Plate: LLM14591 row: n column: 09  
High quality sequence start: 16  
High quality sequence stop: 706.  
Location/Qualifiers  
1..868  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="IMAGE:6959626"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NICHD\_XGC\_Eyel"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.3 kb. Constructed by Life  
Technologies. Note: This is a Xenopus Gene Collection  
(XGC) library."

## FEATURES

source

## ORIGIN

Query Match 6.5%; Score 68.6; DB 14; Length 868;  
Best Local Similarity 45.9%; Pred. No. 0.01;  
Matches 365; Conservative 0; Mismatches 409; Indels 21; Gaps 3;

16 ATCGTGTGAATGTGTGCAACCAATAGTGTGATATGCAAGAAATGATA 75  
72 ATCATGCGGAGATCGGCGAGAACCAAGGCGCATGAAATGCCAAGAAATGATC 131  
76 TTAAGGCCAAGAGCGCGGTGTATGCAATTAATTCCAAACATTTAAAGCTGATTA 135  
132 CTAAAGCGAAGGATGGGTGTCAGATGTGCAAGTTTGAGAAAGAACTGAAATAC 191  
136 TTAATTCAGCTATGCACTTAAGCAGATGATCAATAAATAACACAGAGAAATGAA 195  
132 AAGTCAATTAAGAAAGCTCTGAGAGGACATTAAGTCTCAAGCTGTGGG---GAAA 248  
156 TCTCATGAATGACAAAAAGCTTGAATGAAGTGAAGATATATCTCATCTATG 255  
249 ACATATGAGAGACACACAGCGCAGCTGAGATGACCCATGATACCGGAGATACG 308  
256 GAATATGAGTACAGTTTAATTTAGATTTTCTTCTACCCCTTTGACGAAAGCTAT 315  
309 AAGTATGCAAAAGAGTGAATTTATTTACAGATCTGTGATGAGAGAGGCAATG 368  
316 GATTTTATGATCTTTGAAACAAAAAATATGAAATCCCTTCAAGTGAATTTGAT 375  
369 GAATTCCTTCATGAGCTGGGCGTCCCTTTTCAAGGGGTTCAGGGAGACCAATAC 428  
376 TTACGCTATCTTGAATAAATAGCAAGTTCGATCCCTGATTAAGAAATATATATCA 435  
429 TTCAATATTTAAATAAGACTGCACAAAAGGTGCCCC-----ATGATATATCA 479  
436 ACAGAGATGCTACTATGATGATGAATGAACAGCTGTGTTCTATTTTAAATAATAA 495  
480 AGCGCATGAGCTGCTCTGAAACAAATGCGTGTGTTATGATCTGTGAACAAGTAA 538  
486 GTTCCGGTGTGTAATTTATTAATTTATGATGCAATCTGAATATCCAAAGCCCTT 555  
539 -----TCCAACTTTTGTCTTCAAGGCAAGATGCACTTCAAGAGG 590  
556 GATGTAACCTTAATGCTATTAATGATTTGAAAAACACTTCCCTTAAGATTAATAC 615  
591 GAGCTCAAGCTGTGATGATTAAGAGATTTAGTCTATCTTCCAGATATCCCAAT 650  
616 TTCTGATCATTTAGCGGGTTTATGACGATATGCGGGGCTTATGATTAAT 675  
651 TACTGTGTGATGAAGCTGCAATGCTATCTCAAGTGAAGTGTGCTATGAGCCAAA 710  
676 TTATGAAAAACATTTCACTTATGATTAATGATATGCTGCGCCAGATCATTTGCTCA 735  
711 GTGTGAGAGCGCATGTAACACTGACAAACTTGAAGGAATGATTCACCAAGCTT 770  
736 ATGAGACTGATGAAGCTGAACATTTTGTATGAGGCTGAGGTGTGAAAAATCTTTA 795

Thu May 13 11:53:12 2004

us-09-930-440b-7.rst

Page 7

Db 771 CTGAGCCGACGATGATGACAGATATGTCAGCTCCATCAGACTGTGAGAAAAAGCAATG 830  
QY 796 GGTCAAAATAGTAAA 810  
Db 831 GGGCTCAGCTGTCAAA 845

RESULT 9:  
BM095216/c  
LOCUS DEFINITION  
676 bp mRNA linear EST 20-NOV-2001  
E31811.X1 Sugano SJD adult male Danio rerio cDNA clone  
IMAGE:5410364 3 similar to TR:Q9VG74 Q9VG74 CG5232 PROTEIN. ;  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BM095216 GI:17024182  
EST  
Danio rerio (zebrafish)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.

REFERENCE  
AUTHORS  
1 (bases 1 to 676)  
Clark, M., Johnson, S.L., Iehrach, H., Lee, R., Li, F., Marra, M.,  
Boddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Persson, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
Washu zebrafish EST project 1998

TITLE  
JOURNAL  
COMMENT  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: zbrafish@wustl.edu  
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution information can be found through the I.M.A.G.E.  
Consortium/BLMT, send email to: info@image.llnl.gov  
High quality sequence stop: 525.

FEATURES  
SOURCE  
Location/Qualifiers  
1..676

/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:5410364"  
/sex="male"  
/tissue\_type="whole body"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage resistant)"  
/clone\_lib="Sugano SJD adult male"  
/note="Vector: pME18-FL3; Site 1: DraIII (CACCATGTG);  
Site 2: DraIII (CACTGTGTG); 1st strand cDNA was primed  
with an oligo(dT) primer (ATGCGCCCTTTTCTTTTCTTTT);  
double-stranded cDNA was ligated to a DraIII adaptor  
(CTGTGGCTCTGTG), digested and cloned into distinct DraIII  
sites of the pME18-FL3 vector (5' site CACCATGTG, 3' site  
CACCATGTG). XhoI should be used to isolate the cDNA  
insert. Size selection was performed to exclude fragments  
<1.5kb. Library constructed and donated by Dr. Sumio  
Sugano (University of Tokyo Institute of Medical Science).  
Custom primers for sequencing: 5' end primer  
CTTGGCTCTTAAGCTGCG and 3' end primer  
GCACCTGACGCTGACGACA."

Db 675 TCGAGTGCAAGATGCAATATCCATTTGAACTGAGAGCTCAACCTGCGTTATCCAGG 616  
QY 581 ATTGAAAAAACAATCTCCCTTAAGATATACATAGGCTCTCTGATCATTTAGCGGTTT 640  
Db 615 AATATGAAAGAAATTCACAGACATTCATCGGTACTGAAAGTGGATCA 556  
QY 641 ATGACCTATTGCGGGGCGGCTTATGAAATTAATTTAAGAAAAATTTCACTTAG 700  
Db 555 ACATCACTGCGGGGAGTGTCTTGGGAGCCAAAGTTGAGCGGCAATGACCTGG 496  
QY 701 ATTAATCTATGTCTGCGCCAGATCATTTGCGCTCATATAGACCTGATGAATCAATC 760  
Db 495 ATTAGAGCTGGAAGGCAATGATATGACAGCGCTCTGGAACCAAGAGCTGAGC 436  
QY 761 T 761  
Db 435 T 435

RESULT 10  
CF360452/c  
LOCUS DEFINITION  
673 bp mRNA linear EST 25-AUG-2003  
CF360452  
821784 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.  
CF360452 GI:34158171  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Sus scrofa (pig)  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE  
AUTHORS  
1 (bases 1 to 673)  
Smith, T.P.L., Preking, B.A., Ford, J.J., Vallet, J.L., Fox, J.,  
Wise, T.A., Noneman, D.J., Wray, J.B. and Keeler, J.W.  
A second set of porcine ESTs from a pooled-tissue normalized  
library

TITLE  
JOURNAL  
COMMENT  
Unpublished (2003)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross match v0.990329.  
plate: SRG8013 row: K column: 12  
Seq primer: TAGAGGCAAGTCGAGG.  
Location/Qualifiers  
1..673

/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 3P1G"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
Library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

ORIGIN

Query Match 6.4%; Score 67.6; DB 14; Length 673;  
Best Local Similarity 46.6%; Pred. No. 0.017; Indels 3; Gaps 1;  
Matches 251; Conservative 0; Mismatches 285;  
QY 521 TACATTGCAATGATGATATCCAGCGCTTTGAGATGATTAACCTTAATGCTATTATG 580  
Db 560 TCGAGTGCAAGATGCAATATCCAGCGCTTTGAGATGATTAACCTTAATGCTATTATG 501  
QY 581 ATTGAAAAAACAATCTCCCTTAAGATATACATAGGCTCTCTGATCATTTAGCGGTTT 640  
Db 500 AATATGAAAGAAATTCACAGACATTCATCGGTACTGAAAGTGGATCA 441



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:29 ; Search time 81.495 Seconds

(without alignments)  
7211.402 Million cell updates/sec

Title: US-09-930-440B-7

Perfect score: 1059  
Sequences: 1 atgagtaataatataatcgt.....atcgagaaacaaatattcg 1059

Scoring table: IDENTITY NUC  
Gapop 10-0, Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents NA: \*  
1: /cgm2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgm2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
3: /cgm2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
4: /cgm2\_6/ptodata/2/ina/6C\_COMB.seq: \*  
5: /cgm2\_6/ptodata/2/ina/6D\_COMB.seq: \*  
6: /cgm2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	166.2	15.7 1664976	US-08-916-421B-1	Sequence 1, Appl
2	105.2	9.9 11474	US-09-495-406-1	Sequence 1, Appl
3	63	5.9 1080	US-09-516-13A-3	Sequence 3, Appl
4	63	5.9 1230	US-09-620-312D-1035	Sequence 1035, Ap
5	56.8	5.4 7218	US-08-232-463-14	Sequence 14, Appl
6	52.6	5.0 5152	US-10-204-708-73	Sequence 73, Appl
7	48.4	4.6 640681	US-09-790-988-1	Sequence 1, Appl
8	47.8	4.5 640681	US-09-790-988-1	Sequence 1, Appl
9	47.2	4.5 11049	US-10-204-708-21	Sequence 21, Appl
10	45.6	4.3 15016	US-09-601-198-60	Sequence 60, Appl
11	44.6	4.2 522	US-08-936-165A-12	Sequence 12, Appl
12	44.4	4.2 6866	US-10-204-708-20	Sequence 20, Appl
13	44.4	4.2 26000	US-09-843-376-10	Sequence 10, Appl
14	43.8	4.1 2049	US-09-107-512A-845	Sequence 845, App
15	43.2	4.1 658	US-08-998-416-595	Sequence 595, App
16	43.2	4.1 19124	US-08-487-826B-13	Sequence 13, Appl
17	42.2	4.0 984	US-09-107-512A-2527	Sequence 2527, Ap
18	42	4.0 6243	US-09-056-075-1	Sequence 1, Appl
19	42	4.0 580073	US-08-545-528D-1	Sequence 1, Appl
20	41.8	3.9 356	US-08-520-678A-22	Sequence 22, Appl
21	41.8	3.9 1194	US-08-897-126-22	Sequence 22, Appl
22	41.8	3.9 9646	US-08-811-586-1	Sequence 77, Appl
23	41.8	3.9 9646	US-08-811-586-1	Sequence 1, Appl
24	41.8	3.9 9646	US-08-811-586-1	Sequence 1, Appl
25	41.2	3.9 960	US-09-134-001C-2716	Sequence 2716, Ap
26	41.2	3.9 960	US-09-134-001C-2716	Sequence 2716, Ap
27	41.2	3.9 2126	US-08-235-836C-75	Sequence 75, Appl

#### ALIGNMENTS

```

RESULT 1
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g

```

28	41	3.9	1635	4	US-09-601-198-10	Sequence 10, Appl
29	41	3.9	269223	4	US-09-596-002-41	Sequence 41, Appl
30	40.8	3.9	1686	4	US-09-134-001C-1354	Sequence 1354, Ap
31	40.6	3.8	832	4	US-09-621-976-2813	Sequence 2813, Ap
32	40.6	3.8	2463	4	US-09-134-000C-3032	Sequence 3032, Ap
33	40.6	3.8	580073	4	US-08-545-528D-1	Sequence 1, Appl
34	40.4	3.8	13086	4	US-08-956-171E-16	Sequence 16, Appl
35	40.2	3.8	252	4	US-09-540-236-1345	Sequence 1345, Ap
36	40.2	3.8	5152	4	US-10-204-708-74	Sequence 74, Appl
37	40	3.8	1353	4	US-09-601-198-37	Sequence 37, Appl
38	40	3.8	2448	1	US-08-526-964-2	Sequence 2, Appl
39	40	3.8	2448	2	US-08-546-617-2	Sequence 2, Appl
40	40	3.8	2448	3	US-09-031-897-2	Sequence 2, Appl
41	40	3.8	5203	4	US-09-257-770-1	Sequence 1, Appl
42	40	3.8	99500	4	US-09-798-096-10	Sequence 10, Appl
43	39.8	3.8	2355	3	US-09-371-913A-6	Sequence 6, Appl
44	39.8	3.8	2355	4	US-09-967-805-6	Sequence 6, Appl
45	39.8	3.8	6317	4	US-10-204-708-11	Sequence 11, Appl

```

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g

```

```

NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871519)..(871519)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

Query Match 15.7%; Score 166.2; DB 4; Length 1664976;  
 Best Local Similarity 52.8%; Pred. No. 1.5e-28;  
 Matches 438; Conservative 0; Mismatches 373; Indels 18; Gaps 3;

```

Cy 180 CACGAGATTGACCTGAGTGAAGACAAAACCTGGAATGAAGTACGCA 239
Db 465620 CATAGCAAAAATCAGAAATATGACTATTTAAAGTTTGAACCTGCAAGAGCA 465679
Cy 240 TTATCCATCTATGAGATATGAGTCACTTTAAATTGAGATTTTTCACCCCTTT 299
Db 465680 ATTCTATGAACTAAAATGAGAAAATGGAATATGTTATCTCAACACATT 465739
Cy 300 TGACGAGACTCTATGATTTTTCATCTTTGACATCTTTGACAAACAAAATATGGAATCCCTTC 359
Db 465740 AGATTATTAATATGATATATATTAATTAATGAATGTCCTGCAATTTAAATTCCTTC 465799
Cy 360 AGGTGATTTATGATTTACCGTATCTTGAAAAAATGCAAGCTCCGATCCCTGATTA 419

```

```

Db 465800 TGGTGAATTAACCTTTTATCCCTTATTTAGAAAAAGTGCACAAA-----CAGCCAA 465850
Qy 420 GAAATATATCATATGACAGAAATGGCTACTATTGATAGATATAAACAGTCTGTTTAT 479
Db 465851 GCGGTGATTTTATTTACAGAAATCTGATATTGGGAAATTTGGGAGACATTAAGT 465910
Qy 480 TTTTATTAATAATAAAGTCCGTTGGTATATTAACAATTTACATTCGAAATACTGAATA 539
Db 465911 TTTAATAATATATGATGACAGG-----ATATATTTTATGCTATGATTTCAATCTTA 465964
Qy 540 TCCACGCGCTTGGAGATGTAACCTTAATGCTATTATGATTTGAAAAAACCTTCC 599
Db 465965 CCCAACCCCTTATATAGAGATGCAATTAAGCTATTAACCTTGAAGAAATATATTA 466024
Qy 600 TAAGATAACATAGGCTTCTCTGATCATTTCTAGCGGTTTATGACGCTATTTGCGCGGT 659
Db 466025 TATCCCT---GTGGATATTTCTGACCATATGAGAAATACGCGCCAGATAGTTTCTGT 466081
Qy 660 GCGTTATGGAATACCTTTTATTTGAAAAACATTTCTTATGATTAATCTATGTCGCCC 719
Db 466082 TCCCTTAGAGCGGATGTTATTTAGAGACACTTTACCTTAGATTAATAATATGGAAGTCC 466141
Qy 720 AGATCATTTGGCCTCATATAGAACCTGATGAACGTAAACATCTTTGATTTGGGCTCAGGTG 779
Db 466142 TGTATATGCTTTTGTACAGACAGCCGAGAAATTTAGAAATGTTATATACATTAAGTT 466201
Qy 780 TGTGAAAAATCTTTAGGTTCAATATATATATATGTAAGTGTTCACGCTTCAAGAAAGATATTA 839
Db 466202 AGTTAAAAAAATGCTTGAAGAGCGGAAAGCATACCAATGCTTCTGAAAGAGATAT 466261
Qy 840 AATCTACAGAAAGTCTATATAGCTATTAACAGATATTAATAAAGGTGAGCTTTTTC 899
Db 466262 TGTATAGAGAGAAAGATATTTGTGCAAAAGAAATTTAAAGAAAGAAATATCTTAG 466321
Qy 900 AGAAAAAATATATACACAAAGAAAGCTGTATATGATATCACTCCGATGAGTGTATTA 959
Db 466322 TGTATATATATTTTATTTAAAGACCGGGAGAGATTTGAAACAAAGATTTTAGCAT 466381
Qy 960 TTTATTTGGTAAATTTGAGAGACAGACCTTATTTCCATGATTAATAA 1008
Db 466382 AATATTAACAGAAAAATCAAAAAAGCATTAAGAGAGATGATATTAATA 466430

```

## RESULT 2

US-09-495-406-1

Sequence 1, Application US/09495406

Patent No. 6503744

GENERAL INFORMATION:

APPLICANT: Gilbert, Michel

APPLICANT: Wakarchuk, Warren W.

APPLICANT: National Research Council of Canada

TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of

FLUORESCENTLY LABELED CAMPYLOBACTER AND GANGLIOSIDE ANALOGS

FILE REFERENCE: 019633-00011005

CURRENT FILING DATE: 2000-01-31

PRIOR APPLICATION NUMBER: US 60/118,213

PRIOR FILING DATE: 1999-02-01

NUMBER OF SEQ ID NOS: 35

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 11474

TYPE: DNA

ORGANISM: Campylobacter jejuni

FEATURE: OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384

OTHER INFORMATION: including LOS biosynthesis locus

US-09-495-406-1

Query Match 9.9%; Score 105.2; DB 4; Length 11474;  
Best Local Similarity 49.1%; Pred. No. 4.3e-15;  
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

```

Qy 17 TCGTTCGAAATTTGTTGCAACCATATATGATGTTGATATTCAGAAAGATATAT 76
Db 6979 TCGGCTCGCAAAATAGGATTAATATCATATGCGAGTTAGAACTAGTAAATATATGATG 7038
Qy 77 TAAAGCCAAAGAGCGCGGTGTTAATGACAGTAAATTTCCAAAGCTTAAAGCTATAAT 136
Db 7039 ATGCAAGCTTTAGACAGAGTGTCTAGATTTAAGATTCAAACCCACATCTTGAAGATG 7098
Qy 137 TAAATTCAGCTATGACCTTAGGACAGATATCAATTAATAAACAAGAGAAATTAAGT 196
Db 7099 AGATGAG-----TAAGCCGCTAAATAAAGTAATTCCTGTATATGCAAAATAA 7146
Qy 197 CTCAGTTAGAAATACAAAAAGCTTAATGAGTTAGAGATATCTCATCTAATG 256
Db 7147 GCATTTATGAGATTAATGCAAAATATGCTTAGATTTAAAGATGAGCTTACCTTAAG 7206
Qy 257 AATATGACATGCTTAAATTTATGATGTTTCTACCCCTTTGACGAAGACTATATG 316
Db 7207 AATACAGAAATATATGATGTTTATCTTAGACACACCTTTTCTGTGACAGTGA 7266
Qy 317 ATTTTATGATCTTTGAAACAAAAATATGAAAAATCCCTTCAGGTGATTAATG 376
Db 7267 ACCCTTAGAAAGATATGAGATGTTAGTCTTTAAAGATTTGTCAGGTGATATAT 7326
Qy 377 TACGATCTTGAATAAATAAGCCAGCTCCGATCCCTGATAGAAATATCATATCA 436
Db 7327 ATCCCTTATTAACACATAG-----CAGCTTTAAAGCCTATGATAGTACCA 7377
Qy 437 CAGGAATGCTACTATATGATATGATTAATAACAGTCTGTTCTATTTATTAATAATAAG 496
Db 7378 CAGGATGATATGATTTGAAAGATTAATAACCACTGTAATAATCTTTAGACATATAA 7437
Qy 497 TCCGTTGGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 556
Db 7438 TTCCCTTTGTT-----TTAAATGACACACCAATCTTTACCAACCCCGATATC 7488
Qy 557 ATGTAACCTTAATGCTATTAATGATTTGAAAAACACTTCCCTAGAAATTAACATAGCT 616
Db 7489 TTGTAAGTTAAAGCTATGCTTAATTAATAAAGATTTTCT---TGATGCTAGGCT 7545
Qy 617 TCTGTGATCTTTACGCGGTTTATGACGATATGCGGGGCTTATGATTAATCTT 676
Db 7546 TACCGACACACACAGATATCTTGTGTTAGGTGCGGTGACCTGTGCTGTG 7605
Qy 677 TTATGAAAAACATTTCACTTATGATTAATCTATGTCGCGCCAGATCATTTGCTCAA 736
Db 7606 TCGTTGAAGACATTTTACTGATAGTATGATAGATAGAGCGCCGTGATATAGTTGTTCTA 7665
Qy 737 TAGAAGCTGATGAATGAAACATTTTGTATTTGGGGTCAAGGTGTGAAAAATCTTTG 796
Db 7666 TGGATACACAGGCTTTAAAGAGCTTATTTATCAAGGTGACAAATGCTATATGAGG 7725
Qy 797 GTCAATATGATTAAGTGTATGCTTACGATCAGAAAGAGAAATTAATGTAAGCAAGT 856
Db 7726 GAATTAATGAAAGTAAAGAAAGACCTTAGCAGAGCAAGTCAATGATTTGCTTGG 7785
Qy 857 CTATTAATGCTTAAC---AGAGTAAATAAAGGTGAGGTTTTCAGAAAAATATTA 913
Db 7786 CAAGCTATGATGATTAAGATTAATTAATAAAGCGAAGTTTATCTATGATATATTT 7845
Qy 914 CAACAAAGACCTGCG---TAATGATTCAGTCCAGTGAAGTGAATATTAATTTATG 970
Db 7846 GCGTTAAAGACCTGAGCTTGTGATTTAGATTTAGTGAATTTGAAAAATTTTATGCA 7905
Qy 971 AATATGCAAGACAGA 986
Db 7906 AAAAGCATTAAGAGA 7921

```

## RESULT 3

US-09-516-143A-3

Sequence 3, Application US/09516143A

Thu May 13 11:53:11 2004

us-09-930-440b-7.rnt

Page 4

Patent No. 6333182  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Human Glycosylation Enzymes  
FILE REFERENCE: P505PCT  
CURRENT APPLICATION NUMBER: US/09/516,143A  
CURRENT FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/122,409  
PRIOR FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1080  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1077)  
US-09-516-143A-3

Query Match 5.9%; Score 63; DB 4; Length 1080;

Best Local Similarity 47.8%; Pred. No. 1.2e-05; Mismatches 200; Indels 0; Gaps 0;

Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 518 TATTACATTGCAATACGATATATCCAGCCCTTGAGAGTAAACCTTAATGCTATTA 577  
DB TCTTCAGATGTAACGAGCATACCGCTCCAGCTGAGAGCCTCAACTCGGGGTCACT 601  
QY 578 ATGATTTGAAAAAACAATCTCCCTAGATTAACATAGGCTCTCTGATATCTTACGGCT 637  
DB 602 CGGATATGCAAGCTCTTCTTCTGATCCCATAGGGTATCTGGGCAATAAACAGCA 661  
QY 638 TTATGACGATTTATGCGCGGCTGCTTATGAAATTAATTTTGAATTAACCTTCACTT 697  
DB 662 TAGGATATCTGTGGCCGAGTGGCTCTGGGGGCCAAGGTGTGGAAGTCAATTACTT 721  
QY 698 TAGATTAATCTATGCTGCGCCCAATCATTTGGGCTCAATGAACCTGATTAACCTGAAC 757  
DB 722 TGACACAGCTGGAAGGGAGTGAACCACTCGGCTCTGAGCTGAGCAACTGGCG 781  
QY 758 ATCTTTGATTTGGGCTCAGTGTGTGAAATCTTTAGTTCAATATGTAAGTGTGA 817  
DB 782 AGCTGTGCGGTCACTGCTCTGTGAGCGCTGCGCTCCCAACCAAGAGCTGC 841  
QY 818 CAGCTTCAGAAAGGAATTAATCTGAGCAAGAACTGATTAATGCTAAACAGAGA 877  
DB 842 TGCCCTGTGAGATGCGCTGCAATGAGAAGCTGGGCAAGTGTGTGGCCAAAGTGA 901  
QY 878 TAAAAAAGGTGAGGTTTTC 900  
DB 902 TTCGGAAGGCACCACTTCTAACA 924

RESULT 4

US-09-620-312D-1035

Sequence 1035, Application us/09620312D

Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyun  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Weinman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yundong  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei

APPLICANT: John Tillinghast  
APPLICANT: Dymnac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt.fl\_genes Version 1.0  
SEQ ID NO 1035  
LENGTH: 1230  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (119)..(1198)  
US-09-620-312D-1035

Query Match 5.9%; Score 63; DB 4; Length 1230;

Best Local Similarity 47.8%; Pred. No. 1.2e-05; Mismatches 200; Indels 0; Gaps 0;

Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 518 TATTACATTGCAATACGATATATCCAGCCCTTGAGAGTAAACCTTAATGCTATTA 577  
DB 660 TCTTCAGATGTAACGAGCATACCGCTCCAGCTGAGAGCCTCAACTCGGGGTCACT 719  
QY 578 ATGATTTGAAAAAACAATCTCCCTAGATTAACATAGGCTCTCTGATATCTTACGGCT 637  
DB 720 CGGATATGCAAGCTCTTCTTCTGATCCCATAGGGTATCTGGGCAATAAACAGCA 779  
QY 720 TTATGACGATTTATGCGCGGCTGCTTATGAAATTAATTTTGAATTAACCTTCACTT 697  
DB 780 TAGGATATCTGTGGCCGAGTGGCTCTGGGGGCCAAGGTGTGGAAGTCAATTACTT 839  
QY 780 TAGATTAATCTATGCTGCGCCCAATCATTTGGGCTCAATGAACCTGATTAACCTGAAC 757  
DB 840 TGACACAGCTGGAAGGGAGTGAACCACTCGGCTCTGAGCTGAGCAACTGGCG 899  
QY 758 ATCTTTGATTTGGGCTCAGTGTGTGAAATCTTTAGTTCAATATGTAAGTGTGA 817  
DB 900 AGCTGTGCGGTCACTGCTCTGTGAGCGCTGCGCTCCCAACCAAGAGCTGC 959  
QY 900 CAGCTTCAGAAAGGAATTAATCTGAGCAAGAACTGATTAATGCTAAACAGAGA 877  
DB 960 TGCCCTGTGAGATGCGCTGCAATGAGAAGCTGGGCAAGTGTGTGGCCAAAGTGA 1019  
QY 878 TAAAAAAGGTGAGGTTTTC 900  
DB 1020 TTCGGAAGGCACCACTTCTAACA 1042

RESULT 5

US-08-232-463-14/c

Sequence 14, Application us/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.  
APPLICANT: SCHIEFLINGER, F.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:



Thu May 13 11:53:11 2004

us-09-930-440b-7.rn1

Page 5

```
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/232.463
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/935.313
/ FILING DATE:
/ APPLICATION NUMBER: EP 91 114 300.6
/ FILING DATE: 26-AUG-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BENT, Stephen A.
/ REGISTRATION NUMBER: 29,766
/ REFERENCE/DOCKET NUMBER: 30472/114 IMU
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703)836-9300
/ TELEFAX: (703)683-4109
/ TELEX: 899149
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 7218 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ CLONE: PTZgpc-Fls
/ US-08-232-463-14
```

```
Query Match 5.4%; Score 56.8; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. No. 0.0005;
Matches 16; Conservative 184; Mismatches 116; Indels 0; Gaps 0;
```

```
QY 738 AGAAGCTGATGAAGTGAATCTTTGATTTGGGCTGAGGCTGATGTAATAATCTTAG 797
DB 1464 AGTAGTAAAGAGATAGAAATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1405
QY 798 TTCAATATGTAAGTGTACAGCTTCAGAAAGAGATTAATAATGTCAGAAAGATC 857
DB 1404 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1345
QY 858 TATATAGCTAAAGACAGATTAATAAGTGTATTTTCAGAAATAATTAACAAC 917
DB 1344 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1285
QY 918 AAAAAGCCTGTGATATGATCGATCGATGATGATTAATTTTGGTAAATTTGC 977
DB 1284 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1225
QY 978 AGAGCAAGCTTATTCAGATGATTAATAATTCATAGGATTCAGAAATAGAGGCG 1037
DB 1224 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1165
QY 1038 ATATGAGACAAA 1053
DB 1164 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1149
```

RESULT 6  
US-10-204-708-73/c  
Sequence 73, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: BERLIN, Kurt  
APPLICANT: PIEPENROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06

```
/ PRIOR APPLICATION NUMBER: PCT/EP01/03971
/ PRIOR FILING DATE: 2001-04-06
/ PRIOR APPLICATION NUMBER: DE 10019058.8
/ PRIOR FILING DATE: 2000-04-06
/ PRIOR APPLICATION NUMBER: DE 10019173.8
/ PRIOR FILING DATE: 2000-04-07
/ PRIOR APPLICATION NUMBER: DE 10032529.7
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: DE 10043826.1
/ PRIOR FILING DATE: 2000-09-01
/ NUMBER OF SEQ ID NOS: 98
/ SEQ ID NO 73
/ LENGTH: 5152
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
/ US-10-204-708-73
```

```
Query Match 5.0%; Score 52.6; DB 4; Length 5152;
Best Local Similarity 4.74%; Pred. No. 0.0042;
Matches 251; Conservative 0; Mismatches 274; Indels 4; Gaps 3;
```

```
QY 119 CATTTAAGCTGATTAATTAATTCAGCTATTCAGCTTAAGGCGAGGTATCAATAA 178
DB 1116 CATTTAATTAATTAATTAATTCAGCTATTCAGCTTAAGGCGAGGTATCAATAA 1057
QY 179 ACACAGAGAAATTAAGATCTCAGTTAGAAATGACAAAAGCTGAATGAATGATGAG 238
DB 1056 AACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 998
QY 239 ATATTCATCTAATGAAATGAGTCAAGTGAATTAATTAATTAATTAATTAATTA 298
DB 997 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 938
QY 299 TTGACAGAACTCTAATGATTTTGTAGATCTTTGAAACAAAATAATGAAATGCT 358
DB 937 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 879
QY 359 CAGGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 418
DB 878 CACGAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 819
QY 419 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 478
DB 818 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 759
QY 479 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 536
DB 758 ACTTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 699
QY 537 AATTCAGAGCCCTTGAAGATGTAATTAATTAATTAATTAATTAATTAATTAAT 596
DB 698 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 639
QY 597 CCTAAGATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 645
DB 638 ATACAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 590
```

RESULT 7  
US-09-790-988-1  
Sequence 1, Application US/09790988  
Patent No. 6632935  
GENERAL INFORMATION:  
APPLICANT: SHIGEMO, SHUJI  
APPLICANT: MATANABE, HIDEMI  
APPLICANT: HATTORI, MASAHIRA  
APPLICANT: SAKAKI, YOSHIYUKI  
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
FILE REFERENCE: 081356/0159  
CURRENT APPLICATION NUMBER: US/09/790,988  
CURRENT FILING DATE: 2001-02-23

Thu May 13 11:53:11 2004

us-09-930-440b-7.rn1

Page 6

PRIOR APPLICATION NUMBER: JP2000-107160  
PRIOR FILING DATE: 2000-04-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 1  
LENGTH: 640681  
TYPE: DNA  
ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match  
Best Local Similarity 45.7%; DB 4; Length 640681;  
Matches 169; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 266 TCAGTTAAATTTAGATGTTTTCACCCCTTTGACGAGACCTATGATTTTTCAG 325  
DB 50171 TATTCAGAAATATGATTTTTCATTAATATCTATGATTAATAAATGATTTTAAA 90230  
QY 326 CATCTTGAACAAATATGGAATCCCTCAGTGAATATGATTTTACCATC 385  
DB 90231 GATTAATGATTAATAAATATTAATATGATTAATAAATATTTTGGGATTT 90290  
QY 386 TTGAAAAATAGCCAGCTCCGATCCGATTAAGAAATATATCATCAACAGAAATG 445  
DB 90291 TAACTATATAGCATTTTCAACAAAAAATGATTAAGATTTTATTTATTAAT 90350  
QY 446 CATCATTTGATGATTAATAACGCTGCTTTTATTAATAATTAAGTTCCGATG 505  
DB 90351 TTCCCTTTCTTTTAAATTTTTCACCAATTTTAAAGATTAAGTTGTAAGTTTC 90410  
QY 506 GTAATATTAATATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 565  
DB 90411 AAAGTAAATTAATTAATTAATGATTAATGATTAATGATTAATGATTAATG 90470  
QY 566 TTAATGCTTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 625  
DB 90471 TTAATAATTTTATTTTAAAGATTAATGATTAATGATTAATGATTAATGATTAATG 90530  
QY 626 ATCTAGCGG 635  
DB 90531 ATTTTATAG 90540

RESULT 8  
US-09-790-988-1/c  
Sequence 1, Application US/09790988

PATENT INFORMATION:  
PATENT NO. 6632935  
GENERAL INFORMATION:  
APPLICANT: SHIGENOBU, SHUJI  
APPLICANT: MATSUHARA, HIDEKI  
APPLICANT: HATTORI, MASAHIRA  
APPLICANT: SAKAKI, YOSHIYUKI  
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
FILE REFERENCE: 081356/0159  
CURRENT APPLICATION NUMBER: US/09/790,988  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: JP2000-107160  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 1  
LENGTH: 640681  
TYPE: DNA  
ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match  
Best Local Similarity 50.2%; DB 4; Length 640681;  
Matches 118; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 111 ATCCAAATTTAAAGCTGATTAATTAATTTGACCTATGACCTTAAGGAGATATCA 170  
DB 353778 ATGCAAAATGCTGATTTCTTTTAAATCTTCTATTAATTTCAAGAAAAAGATATTT 353719

QY 171 AATTAATAACAGAGAGATTTAGATTCAGTTAGAAATGACAAAAAGCTTGAATGAA 230  
DB 353718 AATTAATAATAAGAGATTTAGATTCAGTTAGAAATGACAAAAAGCTTGAATGAA 353659  
QY 231 GATGAGATTAATTCATCATTAATGAAATGAGCTGATTAATTTAGATTTTTC 290  
DB 353658 TGTATTTTCTTTTGTATTAATTTTATTTTAAATGAGAAAAATTTATGATTTT 353599  
QY 291 TACCCCTTTGACAGAGACTATGATTTTTCATTTTACATTTTGAACAAAAATA 345  
DB 353598 AATTAATATCAAAATGATTTTCAATCTATGTTTATTTATTAATAAATAAATA 353544

RESULT 9  
US-10-204-708-21/c  
Sequence 21, Application US/10204708

PATENT INFORMATION:  
PATENT NO. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: FIEBENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
PRIOR FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 21  
LENGTH: 11049  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-21

Query Match  
Best Local Similarity 44.9%; DB 4; Length 11049;  
Matches 178; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 96 TGTATGAGATTAATTCAGAAATTAAGCTGATTAATTAATTTGAGCTATGACCC 155  
DB 8049 TCTAATAAATCAATTAATCACTATTAATTAATTAATTTTCTATTTAAACCCAC 7990  
QY 156 TAAGCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 215  
DB 7989 TACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7930  
QY 216 AAGCTGAAATGAGTATGATTAATTCATTAATTAATTAATTAATTAATTAATTAAT 275  
DB 7923 CATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7870  
QY 276 TTTAGATTTTCTTACCCCTTTGACGAGACTATGATTTTATTTTATTTTCAATTA 335  
DB 7869 CTAAATTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7810  
QY 336 AAAAAAATTAAGAAATCCCTGAGTATGATTAATTAATTAATTAATTAATTAATTAAT 395  
DB 7809 ACTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7750  
QY 396 ACCCAAGCTTCGATCCCTGATTAAGAAATTAATTAATTAATTAATTAATTAATTAAT 455  
DB 7749 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7690



Thu May 13 11:53:11 2004

us-09-930-440b-7.rn1

Page 8

RESULT 12

US-10-204-708-20/c

Sequence 20, Application US/10204708

Patent No. 6677731

GENERAL INFORMATION:

APPLICANT: OLEX, Alexander

APPLICANT: PIENBERG, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication

TITLE OF INVENTION: by Assessing DNA Methylation

FILE REFERENCE: 5013.1012

CURRENT APPLICATION NUMBER: US/10/204.708

CURRENT FILING DATE: 2003-05-06

PRIOR APPLICATION NUMBER: PCT/EP01/03971

PRIOR FILING DATE: 2001-04-06

PRIOR APPLICATION NUMBER: DE 10019058.8

PRIOR FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: DE 10019173.8

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: DE 10032529.7

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 98

SEQ ID NO 28

LENGTH: 6866

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-204-708-20

Query Match 4.2%; Score 44.4; DB 4; Length 6866;  
Best Local Similarity 48.8%; Pred. No. 0.35;  
Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

108 AAAATTCGAACATTTAAAGTGAATTAATTAATTTGACATTCAGCTAGGACAGATG 167  
948 AAAATTCCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 889  
168 TCAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 227  
888 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 829  
228 GAAGTATGAGATTAATTCATCATCATCATCATCATCATCATCATCATCATCATCAT 287  
828 AAAACATTCACCAAACTATATTAATAAATAAATAAATAAATAAATAAATAAATAA 769  
288 TTCTACCCCTTTGACGAGATTAATTAATTTGACATTCCTTTGAAACAAAATATG 347  
768 CGCTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 709  
348 GAAAT 353  
708 TAAAT 703

RESULT 13

US-09-843-376-10/c

Sequence 10, Application US/09843376

Patent No. 6566132

GENERAL INFORMATION:

APPLICANT: C. Frank Bennett

APPLICANT: Andrew T. Walt

TITLE OF INVENTION: ANTISENSE MODULATION OF INTERFERON GAMMA RECEPTOR 1 EXPRESSION

FILE REFERENCE: RFS-0234

CURRENT APPLICATION NUMBER: US/09/843.376

CURRENT FILING DATE: 2001-04-26

NUMBER OF SEQ ID NOS: 88

SEQ ID NO 10

LENGTH: 26000

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:

US-09-843-376-10

Query Match 4.2%; Score 44.4; DB 4; Length 26000;  
Best Local Similarity 44.3%; Pred. No. 0.48;  
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

116 AAAATTTAAAGCTGATTAATTAATTTGACATTCAGCTAGGACAGATTAATTAATTA 175  
24115 AT 24056  
176 AAAACAGAGAGATTAATTAATTCAGTTAGAAATGACAAAAGCTTGAAATGAAATG 235  
24055 AT 23996  
236 AGATTAATTCATCTTAATGAAATGAGTCACTTAATTAATTAATTAATTTTCTACCC 295  
23995 TAT 23936  
296 CTTTGAAGAGAGCTCTATGATTTTGGATCTTTGACATCTTTGAAACAAAATATGAAATCC 355  
23935 AACTAT 23876  
356 CTTCAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 415  
23875 TAT 23816  
416 AT 475  
23815 TAT 23756  
476 CTAATTTTAT 521  
23755 AT 23710

RESULT 14

US-09-107-532A-845

Sequence 845, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESS: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107.532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Axinello, Pamela Denike

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-5007

Thu May 13 11:53:11 2004

us-09-930-440b-7.rn1

Page 9

TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 845:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2049 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1..2049  
SEQUENCE DESCRIPTION: SEQ ID NO: 845:  
US-09-107-532A-845

Query March 4.1%; Score 43.8; DB 4; Length 2049;  
Best Local Similarity 50.2%; Pred. No. 0.35; 107; Indels 0; Gaps 0;  
Matches 108; Conservative 0; Mismatches 107; Indels 0; Gaps 0;  
QY 48 TAGTGTGATTCAGAGAGAAATGATTTAAAGCCAGCGTGTATGAGT 107  
DB 1329 TCGGATTCGATTTAAAAAATCGATTTACAGAGAAATGATGTGTTATTTCTAC 1388  
QY 108 AAATTCGAACATTTAAAGCTGATTAATTTAGCTATTCAGCTTAAGCAGAGTA 167  
DB 1389 ATCTATTTTACGAGATACGCTGAGAAATATATTTAGCTCCCGCTACTGCTAGAGA 1448  
QY 168 TCAATTAATAAACAAGAGATTTAGATCTCAGTTGAATGACAAAAAGCTTGAAT 227  
DB 1449 TGAAGTCAGAGCTTAAGAAGCTTTTCGATATTTGAGAGAAAAAAGCAAAAAAT 1508  
QY 228 GAAGTATGACGATTTATCTCCATCTTAATGGAATATG 262  
DB 1509 TACGTATTAACGACATTAACGATGATGATAGAG 1543

RESULT 15  
US-08-998-416-595/C  
Sequence 595, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/R/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 595:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 658 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PA61408BP  
US-08-998-416-595

Query March 4.1%; Score 43.2; DB 3; Length 658;  
Best Local Similarity 43.0%; Pred. No. 0.36; 278; Indels 0; Gaps 0;  
Matches 210; Conservative 0; Mismatches 278; Indels 0; Gaps 0;  
QY 105 AGTAATATTCGAACATTTAAAGCTGATTAATTTAGCTATTCAGCTATTCAGCTTAAGCAG 164  
DB 652 AATTAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAA 593  
QY 165 GTATCAATTAATAAACAAGAGAAATTAAGATTCGTTAGAAATGACAAAAAGCTTGA 224  
DB 592 AATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 533  
QY 225 AATGAATATGAGATTAATCTCCATCTAATGAAATGACGATCACTTAATTTAGATG 284  
DB 532 GAATAGCATATATATATATGTTACAAAATGACATCTATATATATGATTAAGTACACACA 473  
QY 285 TTTTCTACCCCTTTTGAAGAACTCTATGATTTTATAGCTCTTTGAAACAAAAAT 344  
DB 472 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 413  
QY 345 ATGAATAATCCCTTCAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 404  
DB 412 AATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 353  
QY 405 TCCGATCCCTGATTAAGAAATTAATCATATCAACAGGATGCTACTATTTGATGAGTAA 464  
DB 352 AATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 293  
QY 465 ACAGTCTGTTTATTTATTAATAATTAATAAGTCCGTTGATTAATTAATTAATTAATTA 524  
DB 292 ATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 233  
QY 525 TTGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 584  
DB 232 ATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 173  
QY 585 GAAAAAAC 592  
DB 172 AAGCATC 165

Search completed: May 11, 2004, 07:23:09  
Job time: 94.495 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 11:46:35 ; Search time 490.708 Seconds  
(without alignments)  
9776.195 Million cell updates/second

Title: US-09-930-440B-7  
Perfect score: 1059  
Sequence: 1 atgagctaataatatatcgt.....aatggaacaagaattatg 1055

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match 04
	Maximum Match 10

Listing first 45 summaries

```

1:  Published Applications: NA: *
2:  /cgn2_6/prodata/a2/pub/pna/us07_FUBCOMP.seq: *
3:  /cgn2_6/prodata/a2/pub/pna/PCF_NEM_PUB.seq: *
4:  /cgn2_6/prodata/a2/pub/pna/US06_NEM_PUB.seq: *
5:  /cgn2_6/prodata/a2/pub/pna/US06_FUBCOMP.seq: *
6:  /cgn2_6/prodata/a2/pub/pna/US07_NEM_PUB.seq: *
7:  /cgn2_6/prodata/a2/pub/pna/PCFUS_FUBCOMP.seq: *
8:  /cgn2_6/prodata/a2/pub/pna/us08_NEM_PUB.seq: *
9:  /cgn2_6/prodata/a2/pub/pna/us08_FUBCOMP.seq: *
10: /cgn2_6/prodata/a2/pub/pna/US09A_FUBCOMP.seq: *
11: /cgn2_6/prodata/a2/pub/pna/US09_FUBCOMP.seq: *
12: /cgn2_6/prodata/a2/pub/pna/US09_NEM_PUB.seq: *
13: /cgn2_6/prodata/a2/pub/pna/US09_NEM_PUB.seq: *
14: /cgn2_6/prodata/a2/pub/pna/US10A_FUBCOMP.seq: *
15: /cgn2_6/prodata/a2/pub/pna/US10_FUBCOMP.seq: *
16: /cgn2_6/prodata/a2/pub/pna/US10C_FUBCOMP.seq: *
17: /cgn2_6/prodata/a2/pub/pna/US10_NEM_PUB.seq: *
18: /cgn2_6/prodata/a2/pub/pna/us60_NEM_PUB.seq: *
19: /cgn2_6/prodata/a2/pub/pna/us60_FUBCOMP.seq: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1059	100.0	1059	9	US-09-830-440B-7	Sequence 7, Appl
2	326.4	30.8	17276	9	US-09-870-753-83	Sequence 83, Appl
3	326.4	30.8	17276	16	US-09-751-768A-83	Sequence 83, Appl
4	324.8	30.7	25020	10	US-10-193-280-1	Sequence 1, Appl
5	105.2	9.9	11474	9	US-09-816-028A-1	Sequence 1, Appl
6	105.2	9.9	11474	15	US-10-303-161-1	Sequence 1, Appl
7	105.2	9.9	11474	15	US-10-303-161-1	Sequence 1, Appl
8	105.2	9.9	11474	15	US-10-303-128-1	Sequence 1, Appl
9	105.2	9.9	11474	15	US-10-303-134-1	Sequence 1, Appl
10	105.2	9.9	11474	15	US-10-303-162-1	Sequence 1, Appl
11	70.4	6.6	1123	10	US-09-882-227-489	Sequence 489, Appl
12	63	5.9	1080	9	US-09-884-205-3	Sequence 3, Appl
13	63	5.9	1080	9	US-09-830-440B-5	Sequence 3, Appl
14	63	5.9	1230	15	US-10-037-270-1035	Sequence 1035, Appl

C 15	63	1230	16	US-10-117-722-1035	Sequence 1035, App
C 16	5.9	1266	16	US-10-264-137-1192	Sequence 1192, App
C 17	57.2	5.3	1410	US-10-106-698-2649	Sequence 2649, App
C 18	5.3	1023	13	US-10-335-977-2864	Sequence 2864, App
C 19	54.6	5.2	3673778	15 US-10-312-841-1	Sequence 1, Appl
C 20	52.8	5.0	6394	13 US-10-821-613-239	Sequence 239, App
C 21	52.6	5.0	5152	17 US-10-804-708-73	Sequence 73, Appl
C 22	52.6	5.0	5152	17 US-10-240-589-115	Sequence 115, App
C 23	52.2	4.9	942	13 US-10-335-977-2863	Sequence 2863, App
C 24	50.4	4.8	624	13 US-10-027-632-232815	Sequence 232815, App
C 25	50.4	4.8	624	13 US-10-027-632-232816	Sequence 232816, App
C 26	50.4	4.8	624	16 US-10-027-632-232815	Sequence 232815, App
C 27	50.4	4.8	624	16 US-10-027-632-232816	Sequence 232816, App
C 28	49.4	4.7	6061	15 US-10-311-455-114	Sequence 114, App
C 29	48.4	4.6	6406061	9 US-09-790-988-1	Sequence 1, Appl
C 30	48.4	4.5	13326	13 US-10-027-632-123225	Sequence 133225, App
C 31	48	4.5	13326	13 US-10-027-632-123226	Sequence 123226, App
C 32	48	4.5	13326	16 US-10-027-632-123225	Sequence 123225, App
C 33	48	4.5	1326	16 US-10-027-632-123226	Sequence 123226, App
C 34	48	4.5	11650	13 US-10-821-714A-481	Sequence 481, App
C 35	47.8	4.5	572	13 US-10-027-632-194960	Sequence 194960, App
C 36	47.8	4.5	572	16 US-10-027-632-194960	Sequence 194960, App
C 37	47.8	4.5	6406061	9 US-09-790-988-1	Sequence 1, Appl
C 38	47.2	4.5	11049	15 US-10-804-708-21	Sequence 21, Appl
C 39	47.2	4.5	11049	15 US-10-311-455-641	Sequence 641, App
C 40	47.2	4.5	11049	15 US-10-240-589-27	Sequence 27, Appl
C 41	47	4.4	21354	13 US-10-821-714A-512	Sequence 512, App
C 42	46.6	4.4	1136	9 US-09-938-842A-3011	Sequence 3011, App
C 43	46.6	4.4	1136	11 US-09-938-842A-3011	Sequence 3011, App
C 44	46.6	4.4	9504	15 US-10-240-653-281	Sequence 281, App
C 45	45.8	4.3	3673778	15 US-10-312-841-2	Sequence 2, Appl

## ALIGNMENTS

RESULT:  
US-09-930-440B-7  
; Sequence 7, Application US/09930440B  
; Patent No. US20020142386A1

22

;; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways

CURRENT APPLICATION NUMBER: US/09/930,440E

CURRENT FILING DATE: 2001-08-16

PRIOR FILING DATE: 2000-08-25

PRIOR FILING DATE: 2000-03-01

PRIOR FILING DATE: 1999-12-08

PRIOR APPLICATION NUMBER: 60/122,582  
PRIOR FILING DATE: 1999-03-03

NUMBER OF SEQ ID NOS: 8

SEQ ID NO 7

LENGTH: 1059  
TYPE: DNA

ORGANISM: Escherichia coli  
 STRAIN: 8639  
 MEDIA: FERTILIZER

NAME/KEY: CDS  
LOCATION: (1) (10000)

US-09-930-440B-7

Query Match	Score
100.0%	Score

```

Best Local Similarity 100.0%; Pred
Matches 1059: Conservative 0; Min

```

1 ATGAGTAATATATATCGTTGCTG

1

Thu May 13 11:53:12 2004

us-09-930-440b-7.rmp

Page 2

QY 61 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTAAATGAGTAAATTCGAAACA 120  
DB 61 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTAAATGAGTAAATTCGAAACA 120  
QY 121 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAAAATAAAAC 180  
DB 121 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAAAATAAAAC 180  
QY 181 AAGAGAAATTAAGATCTCAGTTAGAAAGACAAAGAAAGCTTGAATGAAGATGACAT 240  
DB 181 AAGAGAAATTAAGATCTCAGTTAGAAAGACAAAGAAAGCTTGAATGAAGATGACAT 240  
QY 241 TATTCATCATATGAGATATGAGATGAGTAAATTTAGATGATTTTCTACCCCTTT 300  
DB 241 TATTCATCATATGAGATATGAGATGAGTAAATTTAGATGATTTTCTACCCCTTT 300  
QY 301 GACGAGACTCTATGATATTTTGAATCTTTGAAACAAATTAAGAAATTCCTTCA 360  
DB 301 GACGAGACTCTATGATATTTTGAATCTTTGAAACAAATTAAGAAATTCCTTCA 360  
QY 361 GGTAGATTAATGATATTCAGCTATTCGAAAGAAATGACCAAGCTTCCATCCCTGATAG 420  
DB 361 GGTAGATTAATGATATTCAGCTATTCGAAAGAAATGACCAAGCTTCCATCCCTGATAG 420  
QY 421 AAATATATCATATCAACAGAAATGCTACTATGATGATGATGATGATGATGATGATGAT 480  
DB 421 AAATATATCATATCAACAGAAATGCTACTATGATGATGATGATGATGATGATGATGAT 480  
QY 481 TTTATATATTAATTAAGTTCGGTGTGAATTAATTAATTAATTAATTAATTAATTAATTA 540  
DB 481 TTTATATATTAATTAAGTTCGGTGTGAATTAATTAATTAATTAATTAATTAATTAATTA 540  
QY 541 CCAAGGCTTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 541 CCAAGGCTTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
QY 601 AAGATATCAATGAGCTTCTCTGATCATCTGAGCGGTTTATGACATGATGAGCGGAGT 660  
DB 601 AAGATATCAATGAGCTTCTCTGATCATCTGAGCGGTTTATGACATGATGAGCGGAGT 660  
QY 661 CATTATGATATCTTTATTTGAAAGAAATTCATCTTATGATGATGATGATGATGATGATGAT 720  
DB 661 CATTATGATATCTTTATTTGAAAGAAATTCATCTTATGATGATGATGATGATGATGATGAT 720  
QY 721 GATCATTTGGCTCAATGAGAACTGATGAACTGATGAACTGATGAACTGATGAACTGATGAACT 780  
DB 721 GATCATTTGGCTCAATGAGAACTGATGAACTGATGAACTGATGAACTGATGAACTGATGAACT 780  
QY 781 GTTGAAGAAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
DB 781 GTTGAAGAAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
QY 841 ATCTGATCAAGAAAGTCTATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
DB 841 ATCTGATCAAGAAAGTCTATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
QY 901 GAAAAAATTAATTAAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
DB 901 GAAAAAATTAATTAAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
QY 961 TTTATGATTAATTAATTAAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
DB 961 TTTATGATTAATTAATTAAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
QY 1021 TTTGAAAAATGAGGAGAAATTAAGAAAGAAATTTATG 1059  
DB 1021 TTTGAAAAATGAGGAGAAATTAAGAAAGAAATTTATG 1059

RESULT 2  
US-09-870-759-83  
Sequence 83, Application US/09870759  
Patent No. US20020177551A1

GENERAL INFORMATION:  
APPLICANT: TERMAN, David S  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
FILE REFERENCE: 870759  
CURRENT APPLICATION NUMBER: US/09/870,759  
CURRENT FILING DATE: 2002-01-14  
PRIOR APPLICATION NUMBER: US 60/208,128  
PRIOR FILING DATE: 2000-05-30  
NUMBER OF SEQ ID NOS: 166  
SOFTWARE: Patent version 3.1  
SEQ ID NO: 83  
LENGTH: 17276  
TYPE: DNA  
ORGANISM: Streptococcus agalactiae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (7062)..(8207)  
OTHER INFORMATION:  
US-09-870-759-83  
Query Match 30.8%; Score 326.4; DB 9; Length 17276;  
Best Local Similarity 58.3%; Pred. No. 1.7e-58;  
Matches 617; Conservative 0; Mismatches 426; Indels 15; Gaps 2;  
1 ATGAGTAAATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60  
DB 12584 ATGAGTAAATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12643  
QY 61 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTAAATGAGTAAATTCGAAACA 120  
DB 12644 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTAAATGAGTAAATTCGAAACT 12703  
QY 121 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAAAATAAAAC 180  
DB 12704 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAAAATAAAAC 12763  
QY 181 AAGAGAAATTAAGATCTCAGTTAGAAAGACAAAGAAAGCTTGAATGAAGATGACAT 240  
DB 12764 AAGAGAAATTAAGATCTCAGTTAGAAAGACAAAGAAAGCTTGAATGAAGATGACAT 12823  
QY 241 TATTCATCATATGAGATATGAGATGAGTAAATTTAGATGATTTTCTACCCCTTT 300  
DB 12824 TATTCATCATATGAGATATGAGATGAGTAAATTTAGATGATTTTCTACCCCTTT 12883  
QY 301 GACGAGACTCTATGATATTTTGAATCTTTGAAACAAATTAAGAAATTCCTTCA 360  
DB 12884 GACGAGACTCTATGATATTTTGAATCTTTGAAACAAATTAAGAAATTCCTTCA 12943  
QY 361 GGTAGATTAATGATATTCAGCTATTCGAAAGAAATGACCAAGCTTCCATCCCTGATAG 420  
DB 12944 GGTAGATTAATGATATTCAGCTATTCGAAAGAAATGACCAAGCTTCCATCCCTGATAG 12994  
QY 421 AAATATATCATATCAACAGAAATGCTACTATGATGATGATGATGATGATGATGATGATGAT 480  
DB 12995 AAATATATCATATCAACAGAAATGCTACTATGATGATGATGATGATGATGATGATGATGAT 13054  
QY 481 TTTATTAATTAATTAAGTTCGGTGTGAATTAATTAATTAATTAATTAATTAATTAATTA 540  
DB 13055 TTTATTAATTAATTAAGTTCGGTGTGAATTAATTAATTAATTAATTAATTAATTAATTA 13108  
QY 541 CCAAGGCTTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 13109 CCAAGGCTTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13168  
QY 601 AAGATATCAATGAGCTTCTCTGATCATCTGAGCGGTTTATGACAGCTATTCGGGAGT 660  
DB 13169 AAGATATCAATGAGCTTCTCTGATCATCTGAGCGGTTTATGACAGCTATTCGGGAGT 13228  
QY 661 CATTATGATATCTTTATTTGAAAGAAATTCATCTTATGATGATGATGATGATGATGATGAT 720  
DB 13229 CATTATGATATCTTTATTTGAAAGAAATTCATCTTATGATGATGATGATGATGATGATGAT 13288  
QY 721 GATCATTTGGCTCAATGAGAACTGATGAACTGATGAACTGATGAACTGATGAACTGATGAACT 780

DB 13289 GATGATTAAGCCAGTGCAGACCCGATTTTACGCTTGTAGTTAAAGGGGTTCCGATT 13348  
QY 781 GTTGAAAAATCTTTAGGTTCAATAGTAAAGTGTACAGTTCCAGAAAGAAATTA 840  
DB 13349 GTTGAAACAGGCTTGTAGTTGAAAAATCCAGATCCAGTGAAGAAAAATAG 13408  
QY 841 ATCTGACCAAGAAAGTCTATTATAGCTAAACAGAGATTAATAAGGTGAGTTTTC 900  
DB 13409 ATTGTCCTGTATATCAGTGTGTTTAAACCAATTAATAAGGATATTATTC 13468  
QY 901 GAAAAAAATATCAACAAAAAGCTGTATAGTATCAGTCCGATGAGTGTATAT 960  
DB 13469 ATAGAAAAATATACGCTGAGAACGCCAGGTATGTATTCCTATGAACTGTGTATG 13528  
QY 961 TTATGGGTAAATATGACAGCAAGACTTATTCAGATGAATTAATATTCATAGCG 1020  
DB 13529 ATCTTGACAGAAAGGCCAGATGATTTCCAGAGAGATGAATTTTCGTGATCA 13588  
QY 1021 TTCAAAAAATCAGGGGGAATTAATAGAACAAAAATTAAT 1058  
DB 13589 TTGAAATCAATTGCCCCGAGTATTAATACACTTAAT 13626

RESULT 3  
US-09-751-708A-83  
Sequence 83, Application US/09751708A  
Publication No. US20030157113A1  
GENERAL INFORMATION:  
APPLICANT: TERMAN, David S  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
FILE REFERENCE: 751708  
CURRENT APPLICATION NUMBER: US/09/751,708A  
PRIORITY FILING DATE: 2002-10-15  
PRIORITY APPLICATION NUMBER: US 60/173,371  
PRIORITY FILING DATE: 1999-12-28  
NUMBER OF SEQ ID NOS: 166  
SOFTWARE: Patent version 3.1  
SEQ ID NO 83  
LENGTH: 17276  
TYPE: DNA  
ORGANISM: Streptococcus agalactiae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (77062)..(8207)  
OTHER INFORMATION:  
US-09-751-708A-83

Query Match 30.8%; Score 326.4; DB 10; Length 17276;  
Best Local Similarity 58.3%; Pred. No. 1.7e-58;  
Matches 617; Conservative 0; Mismatches 426; Indels 15; Gaps 2;  
QY 1 ATGAGTATATATATATGCTGCTGAATTTGTTCCACCACTAATAGTGTGATAT 60  
DB 12584 AAGATTAAGCTTTATATATATGACAGATGCTGTCATCAATTAAGGATTTATCT 12643  
QY 61 GCAAGAAATGATATTAAGCCCAAGAGCGGCTTAATGACAGTAAATTCACAA 120  
DB 12644 GCGAAAAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12703  
QY 121 TTTAAAGTATTAATTAATTTGACCTTAATGACCTTAAGGACAGATATCAATTA 180  
DB 12704 TTTAAAGCTGAGAACTTATTTCTAAATTTCTCCCAAGCTGAATATCAAAAGCA 12763  
QY 181 ACAGAGAAATGATGATGCTGCTGCTGAATTAAGCAAAAAAGCTGAATAGATGAC 240  
DB 12764 ACAGAGAAATGATGATGCTGCTGCTGAATTAAGCAAAAAAGCTGAATAGATG 12823  
QY 241 TATTCATCTAATGAAATATGACGCTTAATTAATTAATTAATTAATTAATTA 300  
DB 12824 TACTAGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 12883  
QY 301 GACGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 360

DB 12884 GATGAAAGCATATAGTCTTAATTTCTACAGATAGCCATTTACAAATTCATCA 12943  
QY 361 GGTGAGTTATTAATTAACCGTATCTTGAAAAATATAGCCAGCTCCGATCCGATAG 420  
DB 12944 GGAATATCTATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 12994  
QY 421 AAAATATCATATCAACAGAAATGCTACTATTTGATGATGATTAATAACATCTGTT 480  
DB 12995 AAGTTATTTCTTCAACGCGTATGCGGTATGGAAGAGATCCATCAAGCGGTAA 13054  
QY 481 TTTATTAATTAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
DB 13055 TTAGCTGAGATGCTACACCG-----ACATTTCTATTTTCAATTTGACAGAG 13108  
QY 541 CCAAGCCCTTTAGAGATGTAACCTTAATGCTATTAATGATTTGAAAAACCTTCCT 600  
DB 13109 CCAACACTTACCCCTCTCTTAATTTAAAGTATTAATCACTTTGAAAGATTA 13168  
QY 601 AAGATTAATTAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
DB 13169 GATTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13228  
QY 661 CTTATGAAATTAATTTATTTGAAAAACATTTCACTTTAGATTAATCTATGCTG 720  
DB 13229 GCAATGGTGCAGAGATTAATTAATAAACCTTACTTATGATGATGAAAGTCC 13288  
QY 721 GATCATTTGCTCTCAATGAACTGATGAACTGAAACCTTTGATTTGGGGTCA 780  
DB 13289 GATCATTAACCCAGTGCACACCTGATTTTAACTGCTGCTTATGATTAAGGGT 13348  
QY 781 GTTGAAAAATCTTTAGTCAATATGATGATGATGATGATGATGATGATGATGAT 840  
DB 13349 GTTGAAACCTTTAGTCAATTTGAAAAATCCAGATCCAGTGAAGAAATTA 13408  
QY 841 ATCTGACCAAGAAATGATTAATGCTTAACACAGATTAATAAGGTGAGTTTTC 900  
DB 13409 ATTGTCCTGTAATCAGTGTGCTTTAAACCAATTTAAAGGCGATTTATTC 13468  
QY 901 GAAAAAAATCAACAAAAAGCTGTATAGTGTACGCTGATGAGTGTATAT 960  
DB 13469 ATAGAAAAATATACGCTGAGAACGCCAGGTATGTATTCCTATGAACTGTGTATG 13528  
QY 961 TTATGGGTAAATATGACAGCAAGACTTATTCAGATGAATTAATATTCATAGCG 1020  
DB 13529 ATCTTGACAGAAAGGCCAGATGATTTCCAGAGAGATGAATTTTCGTGATCA 13588  
QY 1021 TTCAAAAAATCAGGGGGAATTAATAGAACAAAAATTAAT 1058  
DB 13589 TTGAAATCAATTGCCCCGAGTATTAATACACTTAAT 13626

RESULT 4  
US-10-192-280-1  
Sequence 1, Application US/10192280  
Publication No. US20040009574A1  
GENERAL INFORMATION:  
APPLICANT: Datagupta, Nandhushan  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING  
TITLE OF INVENTION: STREPTOCOCCUS AGALACTIAE CAPSULAR POLYSACCHARIDE SYNTHESIS  
TITLE OF INVENTION: GENES  
FILE REFERENCE: 475412001300  
CURRENT APPLICATION NUMBER: US/10/192,280  
PRIORITY FILING DATE: 2002-07-09  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 25020  
TYPE: DNA  
ORGANISM: Streptococcus agalactiae  
US-10-192-280-1





```

Db      7489  TTGTATAGTTAAACCTTATGCTGATTTAAATAAAGATTTTCT---TGATGGTAGGCT  7545
Cy      617  TCTCTGATCAATTTAGACGGGTTTATGACAGCTATGCGCGGTCCTATAGAAATACCT  676
Db      7546  TAAAGGACCAACAACAACATTAATCTTGCGGTGTTTAGGGGCGGTGACCTTGCTGTG  7605
Cy      677  TTATGAAAAAATTTCCTCTAGATAAATCATATGCTGGGCCACATCATTTGGCTCAA  736
Db      7606  TGCCTTGAAGAATTTTACTGATAGTATGCAATAGAGATGGCCCTGATATAGTTGTTCTA  7665
Cy      737  TAGAATCTGATGAACTGAAACATCTTTGTAATGGGGTCAAGGTGTTGAAAAATCTTTAG  796
Db      7666  TGGATATCACAGGCTTTAAAGAGCTTATTTAAACAAGTACCAATATGCTTAATGAGAG  7725
Cy      797  GTTCAATATAGTAATGCTGTTACGCTTGACGAAAGGAATAATTAATCGTACAGAAAGT  856
Db      7726  GAAATATATGAAGATTAATAAAGCAGCTTAACGACAGCAAGTACATTTATTTGCTTTG  7785
Cy      857  CTATTATAGCTTAAAC--AGAGATTAATAAAGGTAGGTTTTTACAAAAAATATTA  913
Db      7786  CAAGCGTATGATGACATTAAGATATTTAATAAAGGAGGAGTTTATCTATGATATATTT  7845
Cy      914  CAACTAAAGACCTTGG---TAAAGTATACGTCGATGAGATGATATTTATTTGGCTA  970
Db      7846  GGGTTTAAAGACCTTGAGCTTGTTGATTAATGACAGCTGAATTTGAAATATTTTAGCA  7905
Cy      971  AAATTGCAGAGCAAGA 986
Db      7906  AAAAAAGCATTAAGAGA 7921

RESULT 6
US-10-303-161-1
: Sequence 1, Application US/10303161
: Publication No. US20030148459A1
: GENERAL INFORMATION:
: APPLICANT: Gilbert, Michel
: APPLICANT: Makarchuk, Warren W.
: APPLICANT: National Research Council of Canada
: TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
: FILE REFERENCE: 019633-000111US
: CURRENT APPLICATION NUMBER: US/10/303,161
: CURRENT FILING DATE: 2002-11-21
: PRIOR APPLICATION NUMBER: US/09/816,028
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: US 60/118,213
: PRIOR FILING DATE: 1999-02-01
: PRIOR APPLICATION NUMBER: US 09/495,406
: PRIOR FILING DATE: 2000-01-31
: NUMBER OF SEQ ID NOS: 49
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 11474
: TYPE: DNA
: ORGANISM: Campylobacter jejuni
: FEATURE:
: OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
: US-10-303-161-1

```

Query Match	9.9%	Score	105.2	DB	15	Length	11474
Best Local Similarity	49.1%	Prod	No.8.5e-12				
Matches	479	Conservative	0	Mismatches	456	Indels	39
						Gaps	6

QY 17 TCGTTGCGCAAAATGGTTCGAACCAATTATGGTAGGTGATTTGCAAGAGAAATGATAT 76  
Db 6979 TCGTCCCTGAATTAGGCATTAATCTATTATGGCAGTTTGAAGTAACTAAATTAATATGCTAG 70338  
QY 77 TAAAGCCCAAGAGCGCCGGTGTAAATGCACTAAATTCGAAACATTTAAAGCTGATAAT 136  
Db 7039 ATGACGCCCTTTAGCAAGGTGCTAAGATTATTAAGCATCAAAACCAACCTGTGAAGTGT 70989

QY	137	TAATTCAGCATTATGCACTTAAGGCAGAGATCAATTAATAAACAACGAGAAATTACAT	196
Db	7099	AGATGAG-----TAAAGCCGCTAAAAGAAAGTAATCCCGGTATAGCAAAAATTA	7146
QY	197	CTCAGTTAAGAAATGACAAAAGCCTGAAATGAAGTATGACGATTATCTCATTTATGG	256
Db	7147	GCATTATATGAGATTATGCAAAAATGTCCTTAGTTATTAAGAAGACTAGCACTTAAG	7206
QY	257	AATATGACATCAGTTTAAATTTAGATGATTTTCTTACCCCTTTGAGGAAGACTATTTG	316
Db	7207	AATACACAGAAAATTTAGCTTGTTTATCTTGACACCTTTTCTCGAGAGTCCAA	7266
QY	317	ATTTTATAGCATCTTGAAACAAAATATGAAATTCCTTCAGGTGACTTATGATTT	376
Db	7267	ACCGCTTGAAGAAATATGGAGTTAGTCTTTTAAAGATTTGTTACGAGTGAATATTT	7326
QY	377	TACCGTATCTTGAATAAATATAGCCAAAGCTCCGATCCCTGTAAAGAAATATCATCAA	436
Db	7327	ATCCGCTTATTAACACATAG-----CAGCCTTTAAAAGCCTATATAGTTAGCA	7377
QY	437	CAGGAATGCGTACATATGATGAGATAAAGCTGCTTTATTTTATATAATTAATAAG	496
Db	7378	CAGGATGAATATGATTTGAAGATATTAACCAATGTAATAATCTTTATAGCAATGAA	7437
QY	497	TTCCGATTGTGAATTTTACAAATATTACATTGCAATATCTGAATTCGAAGCCTTTAGG	556
Db	7438	TTCCCTTTGTT-----TTAATGACACAACCAATCTTTATCCCAACCCCGCATATTC	7488
QY	557	ATGTAACCTTAATGCTATTAAATGATTTGAAAAACACTCCCTAAGATTAACATAGCT	616
Db	7489	TTGTAAGATTAAACGCTATGCTTGATTAATAAAGAAATTTCT---TGATGTAGGCT	7545
QY	617	TCTGTATCATTTCAAGCGGTTTTTATGACCTATTCGCGCGGTGCTTATGGAATACTT	676
Db	7546	TAAAGCAGCACACACAGATTAATCTTCGCTGTATGAGTGGGTGCACTTGCGCTGTG	7605
QY	677	TTAATGAAAAACATTCACCTTATGATTAATCTPAATGTTCGCCCAGATCATTTGGCCCA	736
Db	7606	TGCTTGAAAGACATTTTACTGATATGATGATGAATGAATGGCCCTGATATGTTGTTCTA	7665
QY	737	TAGAACCTGATGAACCTGAACATCTTTGTAATGTGGGCTCAGGTGTGAAAAATCTTTAG	796
Db	7666	TGGATATACAGCGCTTTTAAAGACTTATTATTAACAAATGACCAATATGCTATATAGAG	7725
QY	797	GTTCAATATGATTAAGTGTATACAGCTTCAGAAAGAAAGAAATTAATATGTAACAAGAA	856
Db	7726	GAAATATTAAGAAATTAAGAAAGCAGCTTAAGCAAGCAAGTACCATTTATTTGGCTTGG	7785
QY	857	CTATTATATGCTAAAAC---AGAGTAAAAAAGGTGAGGTTTTTTCAGAAAAAATTTAA	913
Db	7786	CAGGCGTATGCTACATTTAAAGATTTAAAAAAGCGAAAGTTTATCTATGAGATTAATTT	7845
QY	914	CAACAAAAAAGCCTGG---TAATGTATACAGTCCGATGAGAGTGATTAATTTATGGGTA	970
Db	7846	GGGTTAAAGACCTCGACTGTGGATTTAGTGCAGCTGAATTTGAAAAATATTTTAGCA	7905
QY	971	AAATTGACAGACAGA 986	
Db	7906	AAAAAGCATTAAGGA 7921	

RESULT 7  
 US-10-303-118-1  
 : Sequence 1, Application US/10303118  
 : Publication No. US20030157655A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Gilbert, Michel  
 : APPLICANT: Makarchuk, Warren W.  
 : APPLICANT: National Research Council of Canada  
 : TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
 : TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
 : FILE REFERENCE: 019633-00011105

Thu May 13 11:53:12 2004

us-09-930-440b-7.rmpb

Page 6

CURRENT APPLICATION NUMBER: US/10/303,118  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US/09/816,028  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: US 60/118,213  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 09/495,406  
PRIOR FILING DATE: 2000-01-31  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 1  
LENGTH: 11474  
TYPE: DNA  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384  
OTHER INFORMATION: including LOS biosynthesis locus  
US-10-303-118-1

Query Match 9.9%; Score 105.2; DB 15; Length 11474;  
Best Local Similarity 49.1%; Pred. No. 8.5e-12;  
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

17 TCGTGTGCTAAATGTTGACACATATAGTGTGTTGATTTGCAAGAGAAATATAT 76  
DB TCGTGTGCTAAATGTTGACATATATAGTGTGTTGATTTGCAAGAGAAATATAT 7038  
QY 77 TAAAGCCAAAGAGCCGGTGTATATGACATTAATTCACACATTTAAAGCTATAT 136  
DB ATGACGCTTTAGACAGGCTCTAAGATTAAGATCAACCAACCATCTGGAAGATG 7038  
QY 137 TAATTCAGCTATGACCTTACGCAAGATCAATTAAGAAACAGAGGAAATTAAT 136  
DB AGATGAG-----TAAAGCCCTTAAAGAAATTTCTGTAATGCAAAATTA 7146  
QY 197 CTCAGTAAAGATGACAAAGAAAGCTTAATGAAATGAGATTAATTCATCTAATG 256  
DB GCATTTATGAGATTAAGCAAAATGCTTTAGATTTAAGATGAGCTGACCTTAAG 7206  
QY 257 AATATGAGCTGATTAATTAATTAAGATTTTCTTACCCCTTTGACAAAGACTTA 316  
DB AATACAGAGAAATTAAGTCTTGTATCTTACCAACCTTTTCTGTCAGAGTCCAA 7266  
QY 317 ATTATTAAGCATCTTGAACAAAGAAATGAGAAATCCCTGAGTATGATTAAT 376  
DB ACCGCTTAAAGATTAAGGAGTATGCTTTTAAGATGCTTACGAGTATTAAT 7326  
QY 377 TACCGTATCTTGAAGAAATTAAGCAAGCTCCGATCCCTGATTAAGAAATTA 436  
DB ATCGCTTATTAACACATAG-----CAGCTTTAAAGAAAGCTATATGATTA 7377  
QY 437 CAGGAATGCTATGATGAGATTAAGCAAGTGTGTTTATATTTAATTAATTAAG 496  
DB CAGGAATGATTAAGTATTAAGATTAAGCAAGTGTGTTTATATTTAATTAAT 7437  
QY 497 TTCGGTGTGATTAATTAACATTAATTAAGTATTAAGTATTAAGTATTAAG 556  
DB TTCCCTTTGTT-----TTAATGACACCAACCATCTTTACCCACCCGATAT 7488  
QY 557 ATGTAACCTTAATGATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 616  
DB TGTGAATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 7545  
QY 617 TCTGTGATCTATTCAGGCGGTTTATGACGCTATTCGCGGCTTATGATTAAT 676  
DB TAGGCGACCAACAGATTAATCTTCGCGGTTTATGAGTATTCGCGGCTTAT 7605  
QY 677 TTAATGAAAACATTTCACTTAATTAATTAAGTATTAAGTATTAAGTATTAAG 736  
DB TGTCTGAAAACATTTCACTTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 7665  
QY 737 TAGAAGCTGATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 796

DB 7666 TGAATACACAGGCTTTAAAGAGCTTATTAATTAAGTATTAAGTATTAAG 7725  
QY 797 GTTCAATGATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 856  
DB 7726 GAAATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 7785  
QY 857 CTATTAATGCTTAAAC--AGATTAATTAAGTATTAAGTATTAAGTATTAAG 913  
DB CAAGGCTATGACATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 7845  
QY 914 CAACAAAGAGCTG--TATGATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 970  
DB GGGTTAAAGAGCTGATGCTTGTGATTAAGTATTAAGTATTAAGTATTAAG 7905  
QY 971 AATTCAGAGGACAGA 986  
DB 7906 AAAAGCATTAAGAG 7921

RESULT 8

US-10-303-128-1  
Sequence 1, Application US/10303128  
Publication No. US20030157656A1  
GENERAL INFORMATION:  
APPLICANT: Gilbert, Michel  
APPLICANT: Wakarchuk, Warren W.  
TITLE OF INVENTION: National Research Council of Canada  
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
FILE REFERENCE: 019633-0001105  
CURRENT APPLICATION NUMBER: US/10/303,128  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US/09/816,028  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: US 60/118,213  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 09/495,406  
PRIOR FILING DATE: 2000-01-31  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 1  
LENGTH: 11474  
TYPE: DNA  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384  
OTHER INFORMATION: including LOS biosynthesis locus  
US-10-303-128-1

Query Match 9.9%; Score 105.2; DB 15; Length 11474;  
Best Local Similarity 49.1%; Pred. No. 8.5e-12;  
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

17 TCGTGTGCTAAATGTTGACACATATAGTGTGTTGATTTGCAAGAGAAATATAT 76  
DB TCGTGTGCTAAATGACATATATAGTGTGTTGATTTGCAAGAGAAATATAT 7038  
QY 6979 TCGTGTGCTAAATGACATATATAGTGTGTTGATTTGCAAGAGAAATATAT 7038  
QY 77 TAAAGCCAAAGAGCCGGTGTATATGACATTAATTCACACATTTAAAGCTATAT 136  
DB ATGACGCTTTAGACAGGCTCTAAGATTAAGATCAACCAACCATCTGGAAGATG 7098  
QY 137 TAATTCAGCTATGACCTTACGCAAGATCAATTAAGAAACAGAGGAAATTAAT 196  
DB AGATGAG-----TAAAGCCCTTAAAGAAATTTCTGTAATGCAAAATTA 7146  
QY 197 CTCAGTAAAGATGACAAAGAAAGCTTAATGAAATGAGATTAATTCATCTAATG 256  
DB GCATTTATGAGATTAAGCAAAATGCTTTAGATTTAAGATGAGCTGACCTTAAG 7206  
QY 257 AATATGAGCTGATTAATTAATTAAGATTTTCTTACCCCTTTGACAAAGACTTA 316  
DB AATACAGAGAAATTAAGTCTTGTATCTTACCAACCTTTTCTGTCAGAGTCCAA 7266

Thu May 13 11:53:12 2004

us-09-930-440b-7.rnpb

Page 7

Oy	317	ATTTTTGACATCTTTGAAACAAAATAATGAAATATCCCTTCAAGTGTATTAAGAT	376
Db	7267	ACCGCTTGAAAGATATGGAAGCTTAGTGCCTTTTAAGATTGCTTCAGGTAGGTAAATTT	7322
Oy	377	TACCGTATCTTGAAAAATATAGCAAGCTTCGATCCCTGATATGAAAAATATCATATCA	436
Db	7327	ATCCGCTATTAAACACATAG-----CAGCCTTTAAAAAGCCTTATGATAGTAGCA	7377
Oy	437	CAGGAATGCGTACTATGATATGATATGAATAAAGCTGCTGTTTCTATTTTATTAATATTAAG	496
Db	7378	CAGGATATATATGATATGAAAGTATTAATTAACACTGTATTAATTTTATTAAGCAATGAAA	7433
Oy	497	TTCCGGTGTATATTAACAATATTTACATTTGCATATCTGAATATATCCAAAGCCCTTTGAG	556
Db	7438	TTCCCTTTGT-----TTAATAGCACAAACCAATCTTAACCAACCCCGCATTAATC	7488
Oy	557	ATGTAAACCTTAATGCAATTAATGATTTGAAAAAACTCCCTGGAATATCATATAGCT	616
Db	7489	TTGTAAATTAATACGCTATGCTTGATTTAAAAAAGAAATTTCT--TGTATGTAGCT	7544
Oy	617	TCTCGATCATTTCTACGCGGTTTATGACGATATGCGCGGCTGCTTATGGAATTACTT	676
Db	7546	TAAGGACCAACAAACAGATTAATCTTGCGTGTTAAGTGCCTGTGCACCTTGCTGTGG	7605
Oy	677	TTATTGAAAAACATTTCACTTCTTATGATTAATCTATGCTTGCCCAATCATTTGGCCTCA	736
Db	7606	TGCTTGAAGAACAATTTACTGTACTGATATGACATAGAAGTGCCTGATATAGTTTGTCTA	7665
Oy	737	TAGAACCTGATGAATCGAAACAATCTTTGATTTGGGGTCAAGGTGTGTGAAAAATCTTTAG	796
Db	7666	TGGAATACACAGCTTTAAAGAGCTTATTATACAAATGAGCAATGCTATTAATAGAGAG	7722
Oy	797	GTTCAATATGATTAAGTGTATACGCTTCAGAAAGAGATTAATCTGACCAAGAAAGT	856
Db	7726	GAAATTAATGAAGTAAAAAAGGAGGCTAAGCAGAGCAAGATGCACAATTTGATTTGCTTTG	7785
Oy	857	CTATTATGCTTAATAC--AGAGATTAATAAAAAAGTGAGGTTTTCACAAAAAATATTA	913
Db	7786	CAGGTATGCTAGCATTAAGATATTTAAAAAAGGCAAGTTTATCTATGATGATATATTT	7844
Oy	914	CAACAAAAGACCTGG--TAATGATATCAGTCCGATGAGAGTGTATATTTATTTGGGTA	970
Db	7846	GAGGTAAAGACCTGAGACTGTGTGATATTAAGTCAGAGCTGAATTTGAAAATATTTATAGCA	7905
Oy	971	AAATTGCAAGCAACA 986	
Db	7906	AAAAAGCATTAAGGA 7921	

RESULT 9  
 US-10-303-134-1  
 Sequence 1, Application US/10303134  
 Publication No. US20030157657A1  
 GENERAL INFORMATION:  
 APPLICANT: Gilbert, Michel  
 APPLICANT: Makarchuk, Warren W.  
 APPLICANT: National Research Council of Canada  
 TITLE OR INVENTION: Campylodacter Glycosyltransferases for Biosynthesis of  
 TITLE OR INVENTION: Gangliosides and Ganglioside Mimics  
 FILE REFERENCE: 019633-000111US  
 CURRENT APPLICATION NUMBER: US/10/303,134  
 CURRENT FILING DATE: 2002-11-21  
 PRIOR APPLICATION NUMBER: US/09/816,028  
 PRIOR FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: US 60/118,213  
 PRIOR FILING DATE: 1999-02-01  
 PRIOR APPLICATION NUMBER: US 09/495,406  
 PRIOR FILING DATE: 2000-01-31  
 NUMBER OF SEQ ID NOS: 49  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 11474  
 TYPE: DNA

```

; ORGANISM: Campylobacter jejuni
;
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH438
;
; OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-134-1

```

Query Match 9.9%; Score 105.2; DB 15; Length 11474;  
 Best Local Similarity 49.1%; Pred. No. 8.5e-12;  
 Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 5

QY	17	TCGTGCTGAATTCGTTCGACACCATATGATGCTGTGATATTCAGACAGAAATGAT	76
Db	67979	TCGTGCTGAATTCGTTCGACATTCATATATGSCACTTTAGACTGACTTAAATATATGTCG	7033
QY	77	TAAAGCCAAAGAGCCGGGTCTTAATGCAATTAATTCCAACATTTAAAGCTGATAAT	136
QY	137	TAAATTCAGCATTCGACCTTAAGCAGAGTATCAATTAATAAACAACAGAGAAATAGAT	196
Db	7039	ATGCAAGCCTTTAGCACAGGTGCTTAAGATTAAAGCATCAAAACCCACATCGTTGAAGATG	7098
Db	7099	AGATGAG-----TAGGCCGTAAAAAGTAAATTCCTGTATGCAAAAATTA	7144
QY	197	CTCAGTTAGAAATGACAAAAAGCTTGAATGAAGATGACGATATCTCATCTAATG	256
Db	7147	GCATTTATGAGATTATGCAAAAATGTCGTTAGATTATTAAGATAGCTAGCACTTAAG	7200
QY	257	AATATGACATCAGTTTAATTTAATATGTTTTCTACCCCTTTGACGAAGACTCTATG	316
Db	7207	AATACACGAAAAATTAAGTCTGTATTATCTTACACACCTTTTCTCTGCAAGTGC	7266
QY	317	ATTATTAGCATCTTTGAAACAAAAATATGAAATATCCCTCAGTGTAGTTAATTAAT	376
Db	7267	ACCGCTTAGAAGATATGAGATTGTGCTTTTAAGATTGTTCAGGTGATTAATAT	7322
QY	377	TACCGTATCTTGAAAAAATATGCCAAGCTCCGATCCCTGATAGAAAAATATCATATCA	436
Db	7327	ATCCGCTATTTAAACACATAG-----CAGCCTTAAAAACCTTATGATAGTTAGCA	7377
QY	437	CAGGAATGCTACTATTCATGATGAGATAAACAAGTCTGTCTTAATTTTAAATATAATAG	496
Db	7378	CAGGATGAAATGATATGAAAGATATTAACCAACCTGTAATAATCTTATAGCATGAA	7433
QY	497	TTCCGCTGTGTAATATTACATATTTAATCTTCAATGAAATATGCAAGCCCTTTGAG	556
Db	7438	TTCCCTTTGTT-----TTAATGACACACACCAATCTTTACCCACCCGCAATATC	7488
QY	557	ATGTAACCTTAATGCTATTATATGATTTGAAAAAACCTCCCTAGATTAACATGAGCT	616
Db	7489	TTGTAAGATTAAAGCTATGCTTAATTAATAAAGAAATTTCT---TGTAATGTTAGCT	7544
QY	617	TCTCTGATCATTCAGCGGGTTTATGCAAGCTATGCGGGGTCCTTATGGAATACCT	676
Db	7546	TAAAGCAACACAAACAGATTAATCTGTGGTTTATGAGGCGGTGCACTTGGTCTTGTG	7606
QY	677	TTATTTGAAAAACATTCATCTTAATAAATATATGTCGCGCCAGATCATTTGGCTCAA	736
Db	7606	TGCTTGAAGCAATTTTACTGATGATGATGCAATGAAGTGGCCTGATATAGTTGTCTTA	7666
QY	737	TAGAACCTGATGAACCTGAACATCTTTGATATGGGGTCAAGTGTGTTGAAAAATCTT	796
Db	7666	TGATATACACAGGCTTTAAAGAGCTTATTAACAAAGTACCAATATGCTTATATGAGAG	7722
QY	797	GTTCAATATGTAAGTGTATACACTTCAGAAAGGAGAAATTAATGCTACCAAGAAAGT	856
Db	7726	GAAATATATGAAGTAAAAAGCACTAAGCAAGGCAAGTCACAATATATTTGCTCTTG	7788
QY	857	CTATTATATGCTAAATAC---AGAGATAAATAAAGGTGAGTTTTTTCAGAAAAAATATTA	913
Db	7786	CAAGCTATGTCAGCATTTAAAGATATTAATAAAGCGAAGTTTATCTATGATATAATTT	7844
QY	914	CACAAATAAAGCTCTG---TAAATGATACATCCGATGAGATGATATATTTTGGGTA	970
Db	7846	GCGTTAAAGCTCTGAGCTGTGTGAATATGTGCAAGCTAATTTGAAATATTTTAGGCA	7906

Thu May 13 11:53:12 2004

us-09-930-440b-7.rmp

Page 8

QY 971 AAATTGACAGACAGA 986  
DB 7906 AAAAGCATTAGAGA 7921

RESULT 10

US-10-303-162-1  
Sequence 1, Application US/10303162  
Publication No. US20030157658A1  
GENERAL INFORMATION:  
APPLICANT: Gilbert, Michel  
APPLICANT: Makarchuk, Warren W.  
APPLICANT: National Research Council of Canada  
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
FILE REFERENCE: 019633-000111US  
CURRENT APPLICATION NUMBER: US/10/303,162  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US/09/816,028  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: US 60/118,213  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 09/495,406  
PRIOR FILING DATE: 2000-01-31  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 11474  
TYPE: DNA  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384  
OTHER INFORMATION: including LOS biosynthesis locus  
US-10-303-162-1

Query Match 9.9%; Score 105.2; DB 15; Length 11474;  
Best Local Similarity 49.1%; Pred. No. 8.5e-12;  
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

QY 17 TCGTGTCTGAATGTGTGACACATATAGTGTATGTATGCAAGAGAAATGAT 76  
DB 6379 TCGTGTCTGAATGTGTGACACATATAGTGTATGTATGCAAGAGAAATGAT 7038  
QY 77 TAAAGCCAAAGAGCCGGGTGTATGACATTAATCCAAACATTTAAAGCTATAT 136  
DB 7039 ATGACGCTTTAGACACAGGTCTTAAAGATTAAGATCAACCAACATCTTGAAAGT 7098  
QY 137 TAAATTCAGCTATGACCTTAAGCAGATGATCAATTAATAAACAAGAGATTAAT 196  
DB 7099 AGATGAG-----TAAAGCCGCTAAATAAAGATTAATCTGTATGCAAAATAA 7146  
QY 197 CTCGATGAAATGACAAAGAGTTGAATGAATGAGATTAATCTGATCTGATATG 256  
DB 7147 GCATTTATGATTAATGCAAAAGATGCTTTAGATTAAGATTAAGATGACCTTAAG 7206  
QY 257 AATAAGAGTCAGTTTAATTAATTAATGATGTTTCTACCCCTTTAGCAAGACCTATG 316  
DB 7207 AATAAGAGTAATTAATTAATTAATGATGTTTCTACCCCTTTAGCAAGACCTATG 7266  
QY 317 ATTTTATGACATCTTTGAAACAAATAATGAAATATCCCTGAGGATTAATGAT 376  
DB 7267 ACCGCTTAGAATGATGAGATTAATGATGCTTTTAAAGATGCTTGAAGATTAAT 7326  
QY 377 TACCGTATCTTGAATAATAGCAAGCTCCGATCCCTGATAGAAATATATATCA 436  
DB 7327 ATCCGCTATTAACACATAG-----CAGCTTTAAATAAGCTATGATTAAGCA 7377  
QY 437 CAGAAATGCTACTATTAATGATTAATTAAGAGTCTGTTTCTAATTTTAAATAATAG 496  
DB 7378 CAGGAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7437  
QY 497 TTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 556

DB 7438 TTCCCTTTGTT-----TTAATGACACACACACATCTTTAACCACCCGCAATATC 7488  
QY 557 AGGTAACTTATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 616  
DB 7489 TGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7545  
QY 617 TCTGTATCATCTTACGAGGTTTATGACGCTATGCGGCGGTGCTTAATGAAATACT 676  
DB 7546 TAAGCACCACACACACATTAATCTGCTGTTAGTCCGCTGACCTGCTGCTG 7605  
QY 677 TATTAATAAATCACTTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 736  
DB 7606 TGTGTAAGACATTTTACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7665  
QY 737 TAGAACCTGATGAACGTAATCACTTCTTGTATGCGGCTGAGGTGTGTAATAATCTTTAG 796  
DB 7666 TGGATACACAGGCTTTAAAGAGCTTATTAACAAAGTACGCAAAATGCTATATAGAG 7725  
QY 797 GTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 856  
DB 7726 GAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7785  
QY 857 CTTATTAAGCTAAAC---AGAGATTAATAAAGTACGCTTTTTCAGAAATAATTAATTA 913  
DB 7786 CAGGCTAGTACGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7845  
QY 914 CAACAAAGACCTG---TAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 970  
DB 7846 GGGTAAAGACCTGATCTGTTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 7905  
QY 971 AAATTGACAGACAGA 986  
DB 7906 AAAAGCATTAGAGA 7921

RESULT 11

US-09-882-227-489  
Sequence 489, Application US/09882227  
Publication No. US20030158396A1  
GENERAL INFORMATION:  
APPLICANT: Kleanchous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Oomen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in th  
FILE REFERENCE: 06132/047002  
CURRENT APPLICATION NUMBER: US/09/882,227  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 08/902,615  
PRIOR FILING DATE: 1997-07-29  
NUMBER OF SEQ ID NOS: 638  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 489  
LENGTH: 1123  
TYPE: DNA  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (19)...(1038)  
US-09-882-227-489

Query Match 6.6%; Score 70.4; DB 10; Length 1123;  
Best Local Similarity 44.4%; Pred. No. 6.6e-05;  
Matches 434; Conservative 0; Mismatches 526; Indels 18; Gaps 3;

QY 14 ATPTCTGCTGAAATTTGTTGCAACATTAATGATGATGATTAATTAATTAATTAATTAAT 73  
DB 35 AATTTGCTGATTAATGAGCTTAATCAATTAACAGATTTAACTTAGCAAGAAAGGCC 94

Thu May 13 11:53:12 2004

US-09-930-440b-7.rmpb

Page 9

QY 74 TATTAAAGCCAAAGGCGGCTGTATGAGTAATTCGAACATTAAAGCTGATA 133  
DB 95 TTCAATGCCATTAAGGAGGCGGCTGTATGAGTAATTCGAACATTAAAGCTGATA 154  
QY 134 AATTAAATTCAGCTATTCGACCTTAAGGAGTAATTCGAACATTAAAGGAGTAATTCG 193  
DB 155 GCATGACCTTAACTCTAAAGAGATTCCTTCATCATTCAGAGGCACTTTATGGGTAAGA 214  
QY 194 AATTCAGCTAGTAATTCGAACAAAGGCTTGAATTAATGAGCATTAATTCCTCACTCA 253  
DB 215 AAAATTTGTATGAATTTGATCAAAAGGCTTCTACCCCTGAGATGAGCATCTGATTTGT 274  
QY 254 TGAATATGAGCTAGCTATTAATTTAGATGATTTTCTACCCCTTTGACAGAGCTGATA 313  
DB 275 TTGAGTTGGCTTAAGAGCTGATTTAGGCACTTTTACGCTGCTTTAGTTCACAGCTT 334  
QY 314 TTGATTTTTCAGCTATTCGAAACAAAGGCTTGAATTAATTCCTCACTGAGTGTATTGA 373  
DB 335 TAGAGCTTTTGAAGAGCTTAATTCCTCACTGATTAATTCCTGATTTTGAATTCGTTG 394  
QY 374 AATTACCGTATCTGAAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGA 433  
DB 395 AATTGAGCTTGAATTAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGA 445  
QY 434 CAACAGGATGCTATCTATTTGATGATTAATTCAGCTGCTTTTCTATTTTAAATTAATA 493  
DB 446 CTAGGAGTATGCTATCAACAGGATGCTATTCATTCATTTGAGAGAG----- 500  
QY 484 AAGTCCCGGCTGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 553  
DB 501 -AGTAAATTAATTTGACATCACTTGAATTAATTCAGATTAATTCAGATTAATTCAG 559  
QY 554 AGGATTAATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAG 613  
DB 560 AAGAGCTTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAG 616  
QY 614 GCTTCTGATCATTTGAGCGGCTTTATGAGCTATTCGCGGCTGCTTATGAGATAA 673  
DB 617 GCTTGAAGGATCAACGATTCGCTTCTTTGCGGCTTTTGAAGGAGGAGGAGGAGGAG 676  
QY 674 CTTTATTAAGAAACATTTCACTTGAATTAATTCAGATTAATTCAGATTAATTCAG 733  
DB 677 GCATGATTAAGAAACATTTCACTTGAATTAATTCAGATTAATTCAGATTAATTCAG 736  
QY 734 CAATGAACCTGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 793  
DB 737 GCATGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAG 796  
QY 794 TAGGTTCAATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAG 853  
DB 797 TAGGTTCAATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAG 856  
QY 854 ACTTATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 913  
DB 857 GCTTCTTATTTGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 916  
QY 914 CAACAAAGGCTGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 973  
DB 917 AAGCTTTCAGGCTTGAAGGCTTGAATTAATTCAGATTAATTCAGATTAATTCAG 976  
QY 974 TTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 991  
DB 977 AAGGATCAAAATTCCTTAA 994

RESULT 12  
US-09-984-205-3  
Sequence 3, Application US/09984205  
Patent No. US20020137175A1  
GENERAL INFORMATION:  
APPLICANT: Coleman, Timothy A. et al.  
TITLE OF INVENTION: Human Glycosylation Enzymes  
FILE REFERENCE: P50501

CURRENT APPLICATION NUMBER: US/09/984,205  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: PCT/US00/05325  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/516,143  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/122,409  
PRIOR FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1080  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1077)  
US-09-984-205-3  
Query Match 5.9%; Score 63; DB 9; Length 1080;  
Best Local Similarity 47.8%; Pred. No. 0.0024;  
Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;  
QY 518 TATTACATTCGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 577  
DB 542 TCTTGAAGTATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 601  
QY 578 ATGATTTGAAAAACCTTCTCTAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 637  
DB 602 CGAATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 661  
QY 638 TTATGAGCTATTCGCGGCTTCTATGAGATTAATTCAGATTAATTCAGATTAATTCAG 697  
DB 662 TAGGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 721  
QY 698 TAGGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTAATTCAGATTA 757  
DB 722 TGAAGAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 781  
QY 758 ATCTTTGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 817  
DB 782 AGCTGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 841  
QY 818 CAGCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 877  
DB 842 TGCCCTGATGATTCGCTGCAATGAGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 901  
QY 878 TAAAAAAGGTGAGGTTTTCCTCA 900  
DB 902 TTCGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 924

RESULT 13  
US-09-930-440b-5  
Sequence 5, Application US/09930440b  
Patent No. US20020142386A1  
GENERAL INFORMATION:  
APPLICANT: Benenbaugh et al.  
TITLE OF INVENTION: Engineering Intracellular Stajylation Pathways  
FILE REFERENCE: P50592  
CURRENT APPLICATION NUMBER: US/09/930,440b  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: 60/227,579  
PRIOR FILING DATE: 2000-08-25  
PRIOR APPLICATION NUMBER: 09/516,793  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/169,624  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: 60/122,582  
PRIOR FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5

Thu May 13 11:53:12 2004

us-09-930-440b-7.rnpb

Page 10

LENGTH: 1080  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1080)  
US-09-930-440b-5

Query Match  
Best Local Similarity 47.8%; Score 63; DB 9; Length 1080;  
Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 518 TATTACATTGCATTAATTCACAGCCCTTTGAGATGTAAACCTTAATGCTATTA 577  
DB 542 TCTTGACAGTGTACACAGGCAATACCGCTCCAGCCTGAGACGTCAACCTGCGGTCATCT 601  
QY 578 ATGATTTGAAAAACACTTCCCTTAAGATAACATAGGCTTCTGATCAATTTAGCGGCT 637  
DB 602 CGGAATATCGAAGAGCTTTTCTGACATTCCTCATAGGATTTCTGGGCATGAAACAGGCA 661  
QY 638 TTATGACAGTATTCGCGGCTGCTTATGGAATACTTTATGAAAAACATTTCACTT 697  
DB 662 TAGCAATATCTGTGGCGGACAGTGGCTCTGGGGGCCAAGGTGTGAAAGTCAATTAATCT 721  
QY 698 TAGATTAATCTATGCTGTGGCCCATGATTTGGCTTCATAGAACCTGATGAGTGAAC 757  
DB 722 TGGACAAAGACCTGAGAGGAGTACCACTGGCTGTGGAGCTTGAGAACTGAGAGCCG 781  
QY 758 ATCTTTATTTGGGGTCAAGTGTGTGAAAAATCTTTAGGTTCAATAGTAAGTGTTA 817  
DB 782 AGCTGGTGGGCTCAGTGCCTTGTGTGAGGCTGGCCCTCCCAACCAACAGCAGCTGC 841  
QY 818 CAGCTTCAGAAAGAGAAATTAATGCTAGCAAGAAAGTATTAATGCTTAACAGAGA 877  
DB 842 TGCCTGTGAGATGTGCTGCAATGAGAGAGCTGGGCAAGTGTGTGGCCAAAGTGA 901  
QY 878 TAAAAAAGGTGAGGTTTTTTCA 900  
DB 902 TTCCGGAAGGACCATTTCAACA 924

RESULT 14  
US-10-037-270-1035  
Sequence 1035, Application US/10037270  
Publication No. US20030104529A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Weinman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yungding  
APPLICANT: Wang, Dunhui  
APPLICANT: Wang, Zhimei  
APPLICANT: Tillinghast, John  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1 Nucleic Acids and  
FILE REFERENCE: Polypeptides  
CURRENT APPLICATION NUMBER: US/10/037,270  
PRIOR FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104

SOFTWARE: pc\_FL\_genes Version 1.0  
SEQ ID NO 1035  
LENGTH: 1230  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (119)..(1198)  
US-10-037-270-1035

Query Match  
Best Local Similarity 5.9%; Score 63; DB 15; Length 1230;  
Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 518 TATTACATTGCATTAATTCACAGCCCTTTGAGATGTAAACCTTAATGCTATTA 577  
DB 660 TCTTGACAGTGTACACAGGCAATACCGCTCCAGCCTGAGACGTCAACCTGCGGTCATCT 719  
QY 578 ATGATTTGAAAAACACTTCCCTTAAGATAACATAGGCTTCTGATCAATTTAGCGGCT 637  
DB 720 CGGAATATCGAAGAGCTTTTCTGACATTCCTCATAGGATTTCTGGGCATGAAACAGGCA 779  
QY 638 TTATGACAGTATTCGCGGCTGCTTATGGAATACTTTATGAAAAACATTTCACTT 697  
DB 780 TAGCAATATCTGTGGCGGACAGTGGCTCTGGGGGCCAAGGTGTGAAAGTCAATTAATCT 839  
QY 698 TAGATTAATCTATGCTGTGGCCCATGATTTGGCTTCATAGAACCTGATGAGTGAAC 757  
DB 840 TGGACAAAGACCTGAGAGGAGTACCACTGGCTGTGGAGCTTGAGAACTGAGAGCCG 899  
QY 758 ATCTTTATTTGGGGTCAAGTGTGTGAAAAATCTTTAGGTTCAATAGTAAGTGTTA 817  
DB 900 AGCTGGTGGGCTCAGTGCCTTGTGTGAGGCTGGCCCTCCCAACCAACAGCAGCTGC 959  
QY 818 CAGCTTCAGAAAGAGAAATTAATGCTAGCAAGAAAGTATTAATGCTTAACAGAGA 877  
DB 960 TGCCTGTGAGATGTGCTGCAATGAGAGAGCTGGGCAAGTGTGTGGCCAAAGTGA 1019  
QY 878 TAAAAAAGGTGAGGTTTTTTCA 900  
DB 1020 TTCCGGAAGGACCATTTCAACA 1042

RESULT 15  
US-10-117-722-1035  
Sequence 1035, Application US/10117722  
Publication No. US20030219744A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030219744A1 Nucleic Acids and  
FILE REFERENCE: Polypeptides  
CURRENT APPLICATION NUMBER: US/10/117,722  
PRIOR FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pc\_FL\_genes Version 1.0  
SEQ ID NO 1035  
LENGTH: 1230  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (119)..(1198)

US-10-117-722-1035

Query Match 5.9%; Score 63; DB 16; Length 1230;  
Best Local Similarity 47.8%; Pred. No. 0.0025;  
Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY	518	TATTACATTCGAAATCTGAATATCCACGCCCTTGAGAGATTAACCTTAATGCTATTA	577
DB	660	TCCTTCAGTGTACCGGCAATACCGCTCCAGCCTGAGAGCTCACTGCGGTCATCT	719
QY	578	ATGATTTAAAAACACTTCCTTAAGAAATACATAGCCTTCTGTATCATTTAGCGGT	637
DB	720	CGAATATACAGAGCTCTTCCCTGACATTCCTAGGGTATTTGGGATGAAACAGGCA	779
QY	638	TTTATGAGCTATGCGGCGGTGCTTATGGAATTAATTTATGAAAAACATTTCACTT	697
DB	780	TAGCATATCTGTGGCCGAGTGGCTCTGGGGGCCAAGGTGTGAACTGACATTACTT	839
QY	698	TAGATAAATCTATGCTGCGCCCAATCAATTGGCCTCAATAGAACCTGATGAACTGAAAC	757
DB	840	TGACAAAGACCTGGAAGGGAGTGAACACTGCGCTCGCTGAGGCTGAGAACTGCGCG	899
QY	758	ATCTTTGATTTGGGCTCAGGTGTGTGAAAAATCTTTAGCTTCAATAGTAAAGGTTA	817
DB	900	AGCTGTGCGGTCACTGCGTCTTGAGAGCGTGCCTGGGCTCCCAACCAAGCAGCTGC	959
QY	818	CAGCTTCAGAAAGAGAAATAAATCTAGCAAGAACTTATTAAGCTAAACAGAGA	877
DB	960	TGCCCTGTGAGTGGCTGCAATGAGAACTGGGCACTGTGTGTGTCCAAGTGAAGA	1019
QY	878	TAAAAAAGGTGAGGTTTTTCA	900
DB	1020	TTCCGAGGCAACATTTCTAACA	1042

Search completed: May 12, 2004, 12:37:38  
Job time : 501.708 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:49 ; Search time 48.7787 seconds  
(without alignment)  
2004.184 Million cell updates/sec

Title: US-09-930-440B-8  
Perfect score: 1778  
Sequence: 1 MSNIYIVAEICMNGSVDI.....RQDFIPDELIHSEFKNGE 346

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp190s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1778	100.0	346	AAV96102	AAV96102 Escherich
2	1778	100.0	346	AAV96102	AAV96102 Escherich
3	1778	100.0	346	AAV96102	AAV96102 Escherich
4	1016.5	57.2	338	AAV96102	AAV96102 Escherich
5	979.5	55.1	341	AAV96102	AAV96102 Escherich
6	516.5	29.0	359	AAV96102	AAV96102 Escherich
7	514.5	28.9	359	AAV96102	AAV96102 Escherich
8	514.5	28.9	359	AAV96102	AAV96102 Escherich
9	510.5	28.7	359	AAV96102	AAV96102 Escherich
10	510.5	28.7	359	AAV96102	AAV96102 Escherich
11	510.5	28.7	359	AAV96102	AAV96102 Escherich
12	510.5	28.7	359	AAV96102	AAV96102 Escherich
13	497.5	28.0	349	AAV96102	AAV96102 Escherich
14	491.5	27.6	400	AAV96102	AAV96102 Escherich
15	475.5	26.7	346	AAV96102	AAV96102 Escherich
16	475.5	26.7	346	AAV96102	AAV96102 Escherich
17	449.5	25.3	328	AAV96102	AAV96102 Escherich
18	402	22.6	340	AAV96102	AAV96102 Escherich
19	389	21.9	338	AAV96102	AAV96102 Escherich
20	238.5	13.4	201	AAV96102	AAV96102 Escherich
21	224.5	12.6	182	AAV96102	AAV96102 Escherich
22	174.5	9.8	123	AAV96102	AAV96102 Escherich
23	114	6.4	500	AAV96102	AAV96102 Escherich
24	112.5	6.3	1365	AAV96102	AAV96102 Escherich
25	110.5	6.2	1365	AAV96102	AAV96102 Escherich

26	110	6.2	769	6	ABR51668	ABR51668 Protein s
27	109	6.1	601	6	ABU27219	ABU27219 Protein e
28	108.5	6.1	363	6	ABM70260	ABM70260 Phototab
29	107.5	6.0	713	6	ABR53234	ABR53234 Protein s
30	104.5	5.9	436	6	AAV95776	AAV95776 Human pro
31	103.5	5.8	623	6	ABR53138	ABR53138 Protein s
32	103	5.8	653	4	AAV98253	AAV98253 Human hea
33	103	5.8	653	4	AAV98253	AAV98253 Human hea
34	103	5.8	653	4	AAV98253	AAV98253 Human hea
35	103	5.8	950	6	ABU19361	ABU19361 Human mit
36	102	5.7	481	5	ABP73535	ABP73535 Candida a
37	102	5.7	1241	6	ABU24033	ABU24033 Protein e
38	101.5	5.7	650	5	ABP26071	ABP26071 Streptoco
39	101.5	5.7	650	5	ABU46537	ABU46537 Protein e
40	101.5	5.7	841	3	AAV95948	AAV95948 Protein d
41	101	5.7	125	4	ABG17251	ABG17251 Novel hum
42	101	5.7	455	4	ABU52748	ABU52748 Human nuc
43	101	5.7	455	5	AAV47363	AAV47363 Human DEA
44	101	5.7	455	5	AAE20272	AAE20272 Human lun
45	101	5.7	455	6	ADA54078	ADA54078 Human pro

## ALIGNMENTS

RESULT 1  
ID AAV96102 standard; protein; 346 AA.

AAV96102: 19-DEC-2000 (first entry)  
Escherichia coli stialic acid synthetase.  
Stialic acid synthetase; human; sas gene; sialylation; glycoprotein;  
plasmidogen; transferin; chytroctopin; Na<sup>+</sup>,K<sup>+</sup>-ATPase.  
Escherichia coli.  
MO200052135-A2.  
08-SEP-2000.  
01-MAR-2000; 2000MO-US005313.  
02-MAR-1999; 99US-0122582P.  
08-DEC-1999; 99US-0169624P.  
(HUMA-) HUMAN GENOME SCI INC.  
(UTCO-) UNIV JOHNS HOPKINS.  
(UYMY-) UNIV WYOMING.  
Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;  
WPI: 2000-572178/53.  
N-PSDB; AAV50570.  
Recombinant production of sialylated glycoproteins using cells in which  
the expression of enzymes, e.g. stialic acid synthetase, involved in the  
sialylation reaction has been altered.  
Example 6; Page 108; 144pp; English.  
The present sequence is that of Escherichia coli stialic acid synthetase,  
encoded by the nus gene (see AAV50570). The nus gene was used to  
isolate the corresponding human gene (see AAV50569) for stialic acid  
synthetase (see AAV96101). The invention provides methods and  
recombinantly engineered cells for producing glycoproteins having  
sialylated oligosaccharides. The methods involve altering the expression  
of enzymes involved in carbohydrate processing. A claimed cell producing  
sialylated glycoprotein at above endogenous levels expresses at least 1  
(preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme

CC catalyzing the conversion of UDP-GlcNAc to MannNAc, sialic acid  
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid  
 CC transporter at above endogenous levels. A claimed method for manipulating  
 CC glycoprotein in an insect cell comprises enhancing the expression of 1 of  
 CC the above enzymes, and a claimed method for producing sialylated  
 CC glycoproteins involves expressing a heterologous protein (especially  
 CC plasminogen, transferrin, Na<sup>+</sup>/K<sup>+</sup>-ATPase or thyrotropin) in the insect  
 CC cell. Yeast, insect, fungal, plant and bacterial host cells can be  
 CC engineered to produce new forms of sialylated glycoproteins, higher  
 CC concentrations of sialylated glycoproteins and/or elevated concentrations  
 CC of donor substrates (e.g. nucleotide sugars) required for sialylation  
 CC  
 XX  
 SQ Sequence 346 AA;  
 Query Match 100.0%; Score 1778; DB 3; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 4,8e-163;  
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSNTIYVAEIGCNHNSVDIAREMILKAKAGVNAVVFQTFKADKLISAIAPRAEYQIKN 60  
 DB 1 MSNTIYVAEIGCNHNSVDIAREMILKAKAGVNAVVFQTFKADKLISAIAPRAEYQIKN 60  
 QY 61 TGELESQLEMTKLEKMDYDYLHMEYAVSLNDVSTFPDESDIDPLASLKQIKIPS 120  
 DB 61 TGELESQLEMTKLEKMDYDYLHMEYAVSLNDVSTFPDESDIDPLASLKQIKIPS 120  
 QY 121 GELNLNPLYEKLAKLPIDPKKIIISTGMATIDEIKQSVSIFINNKVPVGNITILHCTEY 180  
 DB 121 GELNLNPLYEKLAKLPIDPKKIIISTGMATIDEIKQSVSIFINNKVPVGNITILHCTEY 180  
 QY 181 PTPEDVNINAINDLKXHPKNNIGSDHSGGFYAAIAVPYGITFIEKFTLDKSMGSP 240  
 DB 181 PTPEDVNINAINDLKXHPKNNIGSDHSGGFYAAIAVPYGITFIEKFTLDKSMGSP 240  
 QY 241 DHLASIPDELKHLICIGVRCVEKSLGNSKVVTASEERKNIYAKSIIAKTEIKGEVFS 300  
 DB 241 DHLASIPDELKHLICIGVRCVEKSLGNSKVVTASEERKNIYAKSIIAKTEIKGEVFS 300  
 QY 301 EKNIITKRPNGISPMEMTNLIGKIAEODFIPDELIHSEFRNGE 346  
 DB 301 EKNIITKRPNGISPMEMTNLIGKIAEODFIPDELIHSEFRNGE 346  
 RESULT 2  
 AAB84684  
 ID AAB84684 standard; protein; 346 AA.  
 XX  
 AC AAB84684;  
 XX  
 DT 17-SEP-2001 (first entry)  
 XX  
 DE Amino acid sequence of a bacterial sialic acid synthetase.  
 XX  
 XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;  
 KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;  
 KM sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;  
 KM vaccine.  
 XX  
 OS Escherichia coli.  
 XX  
 XX  
 PN WO200142492-A1.  
 XX  
 PD 14-JUN-2001.  
 XX  
 XX  
 PF 07-DEC-2000; 2000WO-US033136.  
 XX  
 PR 09-DEC-1999; 99US-0169899P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (UYJO) UNIV JOHNS HOPKINS.  
 PA (UTEM) UNIV TEMPLE.  
 PA (UYMY-) UNIV WYOMING.  
 XX

PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;  
 PI Jarvis D;  
 DR WPI, 2001-441575/47.  
 DR N-PSDB; AAH28459.  
 XX  
 PT Cells producing cytidine monophosphate-sialic acid and sialylated  
 PT glycoprotein above endogenous levels for production of vaccines and  
 PT therapeutics.  
 XX  
 PS Example 5; Page 165-166; 182pp; English.  
 XX  
 CC The specification describes a method for manipulating carbohydrate  
 CC processing pathways in cells of interest. The methods are used to  
 CC manipulate multiple pathways involved with the sialylation reaction by  
 CC using recombinant DNA technology and substrate feeding approaches to  
 CC enable the production of sialylated glycoproteins in the cells. The  
 CC sialylation process involves the post-translational addition of the donor  
 CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific  
 CC acceptor carbohydrate. The cells express at least one enzyme, selected  
 CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,  
 CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The  
 CC cells are useful for producing complex sialylated glycoproteins in cells  
 CC of interest, especially insect cells. Glycoproteins containing sialylated  
 CC oligosaccharides are useful as vaccines, therapeutics and diagnostic  
 CC tools. Cells producing complex sialylated glycoproteins are useful for  
 CC enhancing the value of heterologous expression systems and increasing the  
 CC application of heterologous cell expression products as vaccines.  
 CC therapeutics and diagnostic tools as well as increasing the variety of  
 CC heterologous proteins that can be produced and lowering biotechnology  
 CC production costs. The present sequence represents a sialic acid  
 CC synthetase (neub), which is used in the method of the invention  
 CC  
 XX  
 SQ Sequence 346 AA;  
 Query Match 100.0%; Score 1778; DB 4; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 4,8e-163;  
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSNTIYVAEIGCNHNSVDIAREMILKAKAGVNAVVFQTFKADKLISAIAPRAEYQIKN 60  
 DB 1 MSNTIYVAEIGCNHNSVDIAREMILKAKAGVNAVVFQTFKADKLISAIAPRAEYQIKN 60  
 QY 61 TGELESQLEMTKLEKMDYDYLHMEYAVSLNDVSTFPDESDIDPLASLKQIKIPS 120  
 DB 61 TGELESQLEMTKLEKMDYDYLHMEYAVSLNDVSTFPDESDIDPLASLKQIKIPS 120  
 QY 121 GELNLNPLYEKLAKLPIDPKKIIISTGMATIDEIKQSVSIFINNKVPVGNITILHCTEY 180  
 DB 121 GELNLNPLYEKLAKLPIDPKKIIISTGMATIDEIKQSVSIFINNKVPVGNITILHCTEY 180  
 QY 181 PTPEDVNINAINDLKXHPKNNIGSDHSGGFYAAIAVPYGITFIEKFTLDKSMGSP 240  
 DB 181 PTPEDVNINAINDLKXHPKNNIGSDHSGGFYAAIAVPYGITFIEKFTLDKSMGSP 240  
 QY 241 DHLASIPDELKHLICIGVRCVEKSLGNSKVVTASEERKNIYAKSIIAKTEIKGEVFS 300  
 DB 241 DHLASIPDELKHLICIGVRCVEKSLGNSKVVTASEERKNIYAKSIIAKTEIKGEVFS 300  
 QY 301 EKNIITKRPNGISPMEMTNLIGKIAEODFIPDELIHSEFRNGE 346  
 DB 301 EKNIITKRPNGISPMEMTNLIGKIAEODFIPDELIHSEFRNGE 346  
 RESULT 3  
 AA026547  
 ID AA026547 standard; protein; 346 AA.  
 XX  
 AC AA026547;  
 XX  
 DT 06-MAR-2003 (first entry)  
 XX  
 DE Bacterial sialic acid synthetase (Neub) protein.

XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;  
 KM GLNAC-2; epimerase; UDP-GlcNAc; mannose; (Man)Nac; stialic acid;  
 KM synthetase; aldolase; cytidine monophosphate-stialic acid; CMP-SA;  
 KM transporter; stialylated glycoprotein.  
 XX  
 OS Escherichia coli.  
 XX  
 XX US2002142386-A1.  
 XX  
 XX 03-OCT-2002.  
 XX  
 XX 16-AUG-2001; 2001US-00930440.  
 XX  
 XX 02-MAR-1999; 99US-0122582P.  
 XX 08-DEC-1999; 99US-0169624P.  
 XX 25-AUG-2000; 2000US-0227579P.  
 XX  
 PA (BETE/) BETENAUICH M J.  
 PA (LAWR/) LAWRENCE S.  
 PA (LEEY/) LEE Y C.  
 PA (COLE/) COLEMAN T A.  
 XX  
 XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;  
 XX WPI: 2003-102519/09.  
 XX N-PSDB; AAU53994.  
 XX  
 XX Manipulating glycoprotein production in insect cell, involves enhancing  
 PT expression of enzymes involved in carbohydrate processing pathway such as  
 PT N-acetylglucosamine-2 epimerase or stialic acid synthetase.  
 XX  
 XX Disclosure; Fig 35D; 88pp; English.  
 XX  
 XX The invention relates to a novel method for manipulating glycoprotein  
 CC production in an insect cell comprising enhancing expression of an  
 CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one  
 CC catalyzing conversion of UDP-GlcNAc to mannose (Man)Nac, stialic acid  
 CC synthetase, aldolase, cytidine monophosphate-stialic acid (CMP-SA)  
 CC synthetase or CMP-SA transporter, where the expression of each enzyme is  
 CC enhanced to above endogenous levels. The novel method is useful for  
 CC manipulating glycoprotein production in an insect cell. Further methods  
 CC of the invention are useful for producing stialylated glycoprotein. The  
 CC stialylated glycoprotein produced by the above mentioned methods are  
 CC useful as pharmaceutical compositions, vaccines, diagnostics and  
 CC therapeutics. This sequence represents the bacterial stialic acid (NeuB)  
 CC synthetase protein of the invention  
 CC  
 XX Sequence 346 AA;  
 SQ

Query Match 100.0%; Score 1778; DB 6; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 4,8e-163;  
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNIIYVAEICGNHNSVDIAREMILAKKAGVNAVRFQTKAKKLISALPKAEYQIKN 60  
 DB 1 MSNIIYVAEICGNHNSVDIAREMILAKKAGVNAVRFQTKAKKLISALPKAEYQIKN 60  
 QY 1 TGELESQLEMTKKLEMKYDDYILHMEYAVSLNLDVFSTPDESDIDFLASLKQIKWIPPS 120  
 DB 1 TGELESQLEMTKKLEMKYDDYILHMEYAVSLNLDVFSTPDESDIDFLASLKQIKWIPPS 120  
 QY 121 GELINLPYLEKIAKLPIPDKKIISTGMAITDEIKOSYSIFINNKVPGVNTIIHCTEY 180  
 DB 121 GELINLPYLEKIAKLPIPDKKIISTGMAITDEIKOSYSIFINNKVPGVNTIIHCTEY 180  
 QY 181 PTPEDVNAIINLKKHPPKONIGFSDHSGFAAALAAVYGTIPTEKHTLLKSGSGP 240  
 DB 181 PTPEDVNAIINLKKHPPKONIGFSDHSGFAAALAAVYGTIPTEKHTLLKSGSGP 240  
 QY 241 DHLASIEBDELKHLICIGVRCVEKISGNSKVVTASERKNKIIVAKSIIIAKTEIKKGVS 300  
 DB 241 DHLASIEBDELKHLICIGVRCVEKISGNSKVVTASERKNKIIVAKSIIIAKTEIKKGVS 300

QY 301 EKNITTRPNCGISPMEMVNLGKIAEQDEIPDELIHSEPKNOGE 346  
 DB 301 EKNITTRPNCGISPMEMVNLGKIAEQDEIPDELIHSEPKNOGE 346

RESULT 4  
 ID AAY68965 standard; protein; 338 AA.  
 XX  
 XX AAY68965;  
 AC  
 XX  
 DT 30-MAY-2000 (first entry)  
 XX  
 XX Cps2P protein which is involved in stialic acid synthesis.

XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;  
 KM capsular component; antigen; regulation; chain length determination;  
 KM complement-mediated opsonophagocytosis; serotype-specific detection;  
 KM antigen; vaccine; Streptococcal disease; ORF 22; ORF 21; ORF 22; Cps2A;  
 KM Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;  
 KM Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.  
 XX  
 OS Streptococcus suis.  
 XX  
 XX W0200005378-A2.  
 XX  
 XX 03-FEB-2000.  
 PD  
 XX 19-JUL-1999; 99MO-NL000460.  
 PF  
 XX 22-JUL-1998; 98BP-00202465.  
 PR 22-JUL-1998; 98BP-00202467.  
 XX  
 XX (DIEN-) STICHTING DIENST LANDBOUMKUNDIG ONDERZOE.

XX Smith HE;  
 XX  
 XX WPI: 2000-195104/17.  
 DR N-PSDB; AAZ60929.  
 XX  
 XX New nucleic acid containing the capsular gene cluster of Streptococcus  
 PT suis, used for serotype-specific detection and to generate antigens or  
 PT mutants for vaccination.  
 XX  
 PS Disclosure; Fig 3; 144pp; English.

XX The proteins AAY68950-69 are encoded by the capsular gene cluster of  
 CC Streptococcus suis serotype 2. The genes in this cluster are involved in  
 CC polysaccharide biosynthesis of capsular components and antigens. The  
 CC proteins are involved in regulation (CpsA), chain length determination  
 CC (CpsB, CpsC), export (CpsC), and biosynthesis (CpsR, CpsG, CpsH,  
 CC CpsJ, CpsK). The capsule confers bacterium resistance to complement-  
 CC mediated opsonophagocytosis. The gene cluster is used as a source of  
 CC probes and primers for serotype-specific detection of S. suis and is also  
 CC useful for recombinant production of the proteins. The proteins are then  
 CC useful for producing antigens that can be used in vaccines, for  
 CC controlling or eradicating a Streptococcal disease, in humans or animals,  
 CC e.g. against S. suis in pigs

SQ Sequence 338 AA;

Query Match 57.2%; Score 1016.5; DB 3; Length 338;  
 Best Local Similarity 59.2%; Pred. No. 2.2e-89;  
 Matches 202; Conservative 45; Mismatches 89; Indels 5; Gaps 2;

QY 4 IYIVARICGNHNSVDIAREMILAKKAGVNAVRFQTKAKKLISALPKAEYQIKN 63  
 DB 2 VYIIRICGNHNSVDIAREMILAKKAGVNAVRFQTKAKKLISALPKAEYQIKN 61  
 QY 64 LBSQLEMTKKLEMKYDDYILHMEYAVSLNLDVFSTPDESDIDFLASLKQIKWIPPS 123  
 DB 62 SDBQLEMTKKLEMKYDDYILHMEYAVSLNLDVFSTPDESDIDFLASLKQIKWIPPS 121

Qy	184	FEDVNTALNIDLKCHF	PKNNGIGSDHSGCFYALAAVPGYITFIKHTFLDKXNSGPDHL	243
Db	177	YPALNTNVTHTLTKKEE	PNNLTIGYSDSVSEVPIAAMAGELLEKHTFLDNEMEGDHLK	236
Qy	244	ASLEPDELKHL	CGVCEVKSIGSNSKVYTASERKNKIYAKRSIIAKNEIKKGWVPEKN	303
Db	237	ASATPDILAAV	VKGVRIWESLGEFEKEBEVAVRNKIYAKRSIYAKKALANGVFTEN	296
Qy	304	ITTRGPNGSISPMW	NYTLGKIAEOPFIDELIISFEANQ	344
Db	297	ITVGRPNGSISPMW	KYVTLGVSEQDFEEDQWICHSAPENQ	337

RESULT 5  
ABP26810  
ID ABP26810 standard; protein; 341 AA

AC	ABP26810;
XX	
DT	02-JUL-2002 (first entry)

D3 Streptococcus polypeptide SBQ ID NO 2796.

KM Streptococcus; GAS, GBS; group B streptococcus; Streptococcus agalactiae  
 KM Streptococcus; GAS, GBS; group B streptococcus; Streptococcus agalactiae  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002

PF 29-OCT-2001; 2001WO-GB004789.

PR 27-OCT-2000; 2000GB-00026333.

PR 07-MAR-2001; 2001GB-00005640.

PA (CHIR-) CHIRON SPA,

XX  
PT

PI Tettelin H;  
xy

DR WPT; 2002-352536/38.  
DR N-PSDB: A3N67441.

XX  
PT  
New Streets

PT disease caused by  
PT detecting a compound

XX  
PS  
Claim 1: Page 343

xx The invention relates to a protein  
cc

CC (Streptococcus pyogenes), comprising one of 54%

activity. (I), nucleic acids encoding (I), ABNe

CC the treatment or prevention of infection

CC Nucleic acids encoding (1) are used to detect *Streptococcus* in a  
CC biological sample. (1) is used to determine whether a compound binds to  
CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC *Streptococcus* that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (1) may be used to recombinantly produce (1) and may be

CC used in gene therapy. Antibodies to (1) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 341 AA;

Query Match	55.1%	Score	979.5	DB	5	Length	341
Best Local Similarity	56.3%	Prod. No.	8.3e-86				
Matches	192	Conservative	54	Mismatches	90	Indels	5
						Gaps	2

QY 4 IYIVAEIGCHNNGSVIAREKLLKAKEGVNAAYEQTFRADKLISAIAPKAEYOINNTGE 63  
:::||||| :::: | ||:||||| :||| |||||  
DB 2 VYIIAIEIGCHNNGIDINKKWVDVAVSGGDVAKEQTFFRAEKLISKPAKPAAEQKATGT 61

[illegible]

```

OY      124 LNPYLEKAKLPIDDKIIISTGMAITIDEIKÖSVAIFINNKPVGNIITLHONTETPTP 18
        |||||:::||||:::||||:::||||:::|||||
DB      122 TNPYLEKIKG---ÖQKIVILSTGNAWMBEIHQAVILRÖNGST--DISILHCTTEPTPTP 176

```

```
OY      184 FEDVNLNAINDLKHPKNIGFSCHSSGFEYAIAAVPYGTITFIEKHFTLDSKMSGPDHL 24  
       : : : : : : : : : : : : : : : : : :  
Db      177 YPSLTNLNVHTLKQEFKDLTIGYSHSIGSEVPPIAAAMGAEEVIKEHFTLDNTNEGPDK 230
```

```
Oy      244 ASIEPDELKHLICIGVRCVEKSLSGNSKVVTASERKKKIIVAKSIIATIEIKKGVEVPSEKN 300
        |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      237 ASATPDILALAYKSVRIVEQALGREPKIPDVBEKKKIIVAKSVALKPIKKGDYSLEIN 296
```

[illegible]

RESULT 6  
ABB90221  
ID ABB90221 standard; protein; 359 AA

AC AB390221

DT 24-MAY-2002 (first entry)

DE Human polypeptide SEQ ID NO 2597.

KM cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KM antiallergic; hepatotropic; antidiabetic; antinflammatory; anticancer;  
KM valneary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder  
KM neurological disease; infection; human; secreted protein.

Homo sapiens

PN W0200190304-A2

PD 29-NOV-2001

PF 18-MAY-2001; 2001MO-US016450  
 YY

PR 19-MAY-2000; 2000US-0205515P  
XXPA (HUMA-) HUMAN GENOME SCI INC  
YY

PI Birse CE, Rosen CA  
XX

DR WPI; 2002-122018/16  
DR N-2STB: ABL90630

Novel 1405 isolat

PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders.  
XX

PS Claim 11; SEQ ID NO 2597; 2081bp + Sequence Listing; English.

CC The invention relates to novel genes (A1809449-49-BH0085) and proteins  
CC (AAB09040-AAB09444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (anti)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
CC (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

SQ Sequence 359 AA;

Query Match	29.0%	Score 516.5	DB 5	Length 359
Best Local Similarity	36.6%	Pred. No. 5.5e-41		
Matches 124; Conservative	61	Mismatches 139	Indels 15	Gaps 6

[illegible]

XX	RESULT 7
XX	AA039986
XX	ID AAM39986 standard; protein; 359 AA.
XX	AC AAM39986;
XX	DT 22-OCT-2001 (first entry)
XX	DE Human polypeptide SEQ ID NO 3131.
XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX	peripheral nervous system; neuropathy; central nervous system; CNS;
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia.
XX	Homo sapiens.
XX	OS
XX	PN WO200153312-A1.
XX	DT 26-JUL-2001.

XX 26-DEC-2000; 2000MO-US034263.  
PF  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662151.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wenman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou F, Goodrich R, Dimañac RT;  
XX  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI59142.  
XX  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PI as central nervous system injuries.  
XX  
PS Example 4; SEQ ID NO 3131; 10078pp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
CC encoded polypeptides (AA138642-AA1442213) with neurotropic,  
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotides  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombotic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification

**SQ Sequence 359 AA;**

Query Match	28.9%	Score 514.5	DB 4	Length 359
Best Local Similarity	36.6%	Pred. No. 8-66-41		
Matches 124; Conservative	60;	Mismatches 140;	Indels 15;	Gaps 6;

Qy YTAELGCHNHSVPIARPMILKAEDAVANVKOTFADLILSALIPKAEYOIKYNGEL 64  
Db FIIAELGOMHGDVLAKAMITMAEBCADCAKQKSLSEKFKRKMLDRPYTSKHWG- 78  
Qy ESQLEMTKLBMKYDDYLHMEYAVSLNLDVSTPEPDSIDPLASIKOKIKWIPSGELL 124  
Db KTYGEKRRHLESHQYELQRYAEWVGIFPLASGDMEMAEFLHNLNVEFFKVGSGDTN 138  
Qy NLTYELKIKLPIPKKIITISQVATIDEIKOSVIFINNVPVGNITLHCNTEYPTF 184  
Db NPYELTKIKAKGRP---WISISQMSMTMVOYQIV---KELNPNCFQCSAYELQ 192  
Qy 185 EDVNINAINDLKHPKKNIGFSDHSQGYAIAAVPYGITFIEKHEFTLDKSSGPDHIA 244  
Db 193 EDVNLEVISEYOKLPDPIFGSGEFTIALSTAAVAALGKAVLEHHTLLDNTWGSQSHSA 252  
Qy 245 SIEPDELKHLCTGVACVSKSLGSNKKYVTAASERKNQIVAKRSITAKIEIKKGEVSEKNI 304  
Db 253 STEPGELALVSVLVEBALGSPYKQLLPDEMACNEKLGKSVAKYKITEGTLITMDNL 312  
Qy 305 TTK-RPGNGISPMEMYNILGK-----IABODFTPDLIL 336  
Db 313 TVKVGEP-KGYPPEDIFNLVGRKVLVLAEEDDITMEBIV 350

RESULT 8  
 AAB93183  
 ID AAB93183 standard; protein, 359 AA.  
 XX  
 AC AAB93183;  
 XX  
 DT 26-JUN-2001 (first entry)  
 DE Human protein sequence SEQ ID NO:12130.  
 XX  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX Homo sapiens.  
 XX EPI074617-A2.  
 XX 07-FEB-2001.  
 PF 26-JUN-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 PS  
 PS Claim 8; SEQ ID NO 12130; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide which comprises one of the 5602  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification; where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence; where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any special methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 359 AA;  
 XX  
 Query Match 28.9%; Score 514.5; DB 4; Length 359;  
 Best Local Similarity 36.6%; Pred. No. 8.6e-41;  
 Matches 124; Conservative 60; Mismatches 140; Indels 15; Gaps 6;  
 QY 5 YIYAEIGCNGNSVDIAREMILKAKENGUNAVFQFFKADKLISALAPKAEYIKNTGEL 64  
 DB 20 FIIAEIGQNHQGLDVAKRMIRAKCGACCAKFKSELEFKRRKALDRPYTSKRSWG- 78

QY 65 ESQLEMTKTLKMKYDYLHMEYAVSLNDVSTPEDESDIFLASLKOKIWKIPSGELL 124  
 DB 79 KTYGEHKKRLHLEPSHQYRELQRYAEFVCIFFLASGDEMAVFEHLNLPFPKVGSGDNN 138  
 QY 125 NLPYLEKIKAKLPIPDKKIISTGMATIDEIKOSVSFFINNKYIPVGNITILHCNTEPTPE 184  
 DB 139 NFPYLEKIKAKKGRP--WVLSGQSGMDTKQVQYQV--KELNRPFCLOCTSAVFLQF 192  
 QY 185 EDVNLNAINDLKHPKXNIGFSDHSGFYAAIAAVPYGITEFKFTLDKSMGPDHIA 244  
 DB 193 EDVNLNAISEYQKLPFDIPDIGSGHETGIAISVAVALGTKYLERHITLDKTKWKGSDHSA 252  
 QY 245 STEPDELKLCIGYRVVEKSLGNSKVYVASERKKIVARKSLIKTEIKKEGVESEKNI 304  
 DB 253 SLFPGELAEVRSVRLVERALGSPFKQLPCEMACNEKLGKSVAKVLPBGITLMDWL 312  
 QY 305 TTK--SPNGISPMEWYNLQK----IAEODFIPDELI 336  
 DB 313 TVKVGSP-KGYPPEDIFNLVGKKLVTVFEDDTIKRELV 350  
 XX  
 RESULT 9  
 AAY96101  
 ID AAY96101 standard; protein, 359 AA.  
 XX  
 AC AAY96101;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Human sialic acid synthetase.  
 XX  
 KM Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;  
 KM plasminogen, transferrin, thyrotropin, Na+, K+-ATPase.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200052135-A2.  
 XX  
 PD 08-SEP-2000.  
 XX  
 PF 01-MAR-2000; 2000WC-US005313.  
 XX  
 PR 02-MAR-1999; 99US-0122582P.  
 PR 08-DEC-1999; 99US-0169624P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (UYUO) UNIV JOHNS HOPKINS.  
 PA (UYWY-) UNIV WYOMING.  
 XX  
 PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;  
 DR WPI; 2000-572178/53.  
 DR N-PSDB; AAA50569.  
 XX  
 PT Recombinant production of sialylated glycoproteins using cells in which  
 PT the expression of enzymes, e.g. sialic acid synthetase, involved in the  
 PT sialylation reaction has been altered.  
 XX  
 XX Claim 16; Page 105-106; 144pp; English.  
 XX  
 CC The present sequence is that of human sialic acid synthetase (SAS), an  
 CC enzyme that condenses Mann6-6-P or Man6-6-P with PRP to form Neu5Ac and  
 CC KDN phosphates, respectively. The sequence was deduced from SAS cDNA (see  
 CC AA50569). Northern blots indicated ubiquitous transcription of the SAS  
 CC gene in a selection of tissues. The invention provides methods and  
 CC recombinantly engineered cells for producing glycoproteins having  
 CC sialylated oligosaccharides. The methods involve altering the expression  
 CC of enzymes involved in carbohydrate processing. A claimed cell producing  
 CC sialylated glycoprotein at above endogenous levels expresses at least 1  
 CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme  
 CC catalyzing the conversion of UDP-GlcNAc to Mann6c, sialic acid  
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid



[illegible]

Db	79	KTGEHKKRLFFSHQYRELQRYAEVGIFFTAGCDMAVEFLHNLNVEFFRVSGDDTN	138
Qy	125	NLFYELKIAKLPIPDKKIITSGATIDELQKQSVIFLNKVPVGNITTHCNTETPTF	184
Db	139	NPFYLEKTKAKGSP--NWISGGMSQMTQKQVQIV---KELNPNCFCLQCSAAYPLQF	192
Qy	185	EDVNLNAINADLKHPKNNIGPSDHSSEFAAIAAVPYGIFIEIKHFTLDRKSMGPDHLA	244
Db	193	EDVNLKVISBYQKLPDPIDPIGSGHETIAISVAANLQKATLERHITLDRTKKSGDHS	252
Qy	245	SIEPDELKHLICIGVRCEKSLGSSNKKVYTAERNNKTIYAKRSIIIAKTEIKKGEVPEKON	304
Db	253	SLFPGELAEIVASVRLVERALGSPFKQLLPCPMACNENKLGKSVAAKVIPEGITILTWDL	312
Qy	305	TTK--RPGNGISPMETVNLGK----IAEDDFIPDELI	336
Db	313	TVRVGEP-KAYPEDIPLNLGKVLIVAEEDDTIMEELV	350
RESULT 12			
ID	AA026546	AA026546 standard; protein; 359 AA.	
XX	AA026546;		
XX	06-MAR-2003	(first entry)	
XX	Human SA synthetase protein.		
XX	Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;		
XX	GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAc; sialic acid;		
XX	synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;		
XX	transporter; sialylated glycoprotein; human.		
XX	Homo sapiens.		
XX	US2002142386-A1.		
XX	03-OCT-2002.		
XX	16-AUG-2001; 2001US-00930440.		
XX	02-MAR-1999; 99US-0122582P.		
XX	08-DEC-1999; 99US-0169624P.		
XX	25-AUG-2000; 2000US-0227579P.		
PA	(BETR/) BETENBAUGH M J.		
PA	(LAWR/) LAWRENCE S.		
PA	(LEEV/) LEE Y C.		
PA	(COLE/) COLEMAN T A.		
PI	Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;		
DR	WPI; 2003-102519/09.		
DR	N-PSDB; AAL53993.		
PT	Manipulating glycoprotein production in insect cell, involves enhancing		
PT	expression of enzymes involved in carbohydrate processing pathway such as		
PT	N-acetylglucosamine-2 epimerase or sialic acid synthetase.		
PS	Claim 16; Fig 32; 88pp; English.		
CC	The invention relates to a novel method for manipulating glycoprotein		
CC	production in an insect cell comprising enhancing expression of an		
CC	enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one		
CC	catalyzing conversion of UDP-GlcNAc to mannose (Man)NAc, sialic acid		
CC	synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)		
CC	synthetase or CMP-SA transporter, where the expression of each enzyme is		
CC	enhanced to above endogenous levels. The novel method is useful for		
CC	manipulating glycoprotein production in an insect cell. Further methods		
CC	of the invention are useful for producing sialylated glycoprotein. The		
CC	sialylated glycoprotein produced by the above mentioned methods are		



CC useful as pharmaceutical compositions, vaccines, diagnostics and  
 CC therapeutics. This sequence represents the human SA synthetase protein of  
 CC the invention

XX Sequence 359 AA:

Query Match 28.7%; Score 510.5; DB 6; Length 359;  
 Best Local Similarity 36.3%; Pred. No. 2.1e-40;  
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 5 YVAVICGNNGSVDIRRMILKAKKGVNAVKQTRKADKLISALPKAEYQIKNGEL 64  
 DB 20 FIIAIEIQNNHGGDIDVAKRMIRMAKCEGACDCKFKSELEKFKKLERPYTSKHSWG- 78  
 QY 65 ESQLEMTKLEMKDYDLHMEYAVSINLDVSTPDESDIDFLASIKOKIWKIPSGELL 124  
 DB 79 KTYGEHGRHLEFSDQYRELQRYAEYGIFFTLASGDMEMVFLHMLNVPFKXGSDEN 138  
 QY 125 NLPLEKIAKLPIDPKKIIITSGATIDEIKOSVITINNKVPVGNITIIHCNTEYTPF 184  
 DB 139 NFPLEKTAAGKRP---MVISSGQSMDFMKQVQIV---KPLNPNCFLOQTSAYPLQP 192  
 QY 185 EDVNAINIDLKPKPKNNIGFSDHSGFYAALAAVYGTIFIEKHFTLTKSMGPDHLA 244  
 DB 193 EDVNAIRYISYQKLPDIPIGYSGETGIALSVAAVLGAKVLERHITLDTKTKGSDHSA 252  
 QY 245 SIEPDELKHLCTIGRCVEKSLGNSKVVLTASERKNKIIVARKSIIAKTEIKKGEVSEKNT 304  
 DB 253 SLRGEIAELVRSVRLVERALGSPKQLPCEMACNKEKSGSVAAKVIPEGITILLTMDKL 312  
 QY 305 TTK--RPGNGISPMENYILGK----IAQDPIDDELI 336  
 DB 313 TVKVGEP-KAYPPEDITNLVGKVLTVVEEDTMEELV 350

#### RESULT 13

AA49716 ID AAM49716 standard; protein; 349 AA.

AC AAM49716;  
 DE 07-JUN-2002 (first entry)

DE N. meningitidis sialic protein.

XX N-acetylneuraminic acid synthase; sialic acid; malignant cell;  
 XX inflammatory process; cell surface glycoprotein; glycolipid;  
 XX cell growth regulation; cell differentiation; metastasis.

XX Neisseria meningitidis.

XX DE10034586-A1.

XX 07-FEB-2002.

XX 14-JUL-2000; 2000DE-01034586.

XX 14-JUL-2000; 2000DE-01034586.

XX (FEBS/) FEBSNER W.

XX Fessner W, Knorst M;

XX WPI; 2002-29312/34.

XX N-PSDB; ABA99765.

XX High activity recombinant N-acetylneuraminic acid synthase production,  
 XX for use in the preparation of N-acetylneuraminic acid compounds,  
 XX including new derivatives and analogs.

XX Claim 4; Page 20-22; 26pp; German.

XX This invention describes a novel preparation of proteins having N-

CC acetylneuraminic acid synthase activity. The preparation involves  
 CC culturing a prokaryotic host organism transformed by an expression vector  
 CC containing a structural gene encoding proteins with N-acetylneuraminic  
 CC acid synthase activity. The proteins of the invention are useful in the  
 CC production of sialic acids and their derivatives and analogues. Sialic  
 CC acids are components of cell surface glycoproteins and glycolipids, and  
 CC are involved in physiological and pathological recognition processes,  
 CC e.g. in inflammatory processes, regulation of cell growth and  
 CC differentiation, malignant cell dedifferentiation and metastasis or as  
 CC recognition domains of pathological viruses and bacteria. Glycoproteins  
 CC and glycolipids based on non-natural sialic acid derivatives and  
 CC analogues may be useful as metabolically stable therapeutic agents or as  
 CC fluorescently labeled analytical/diagnostic agents. The proteins obtained  
 CC by the present method are more accessible than prior art analogues, and  
 CC have significantly higher specific activity, markedly wider substrate  
 CC tolerance and superior stability. This sequence represents the Neisseria  
 CC meningitidis sialic gene described in the disclosure of the invention

XX Sequence 349 AA:

Query Match 28.0%; Score 497.5; DB 5; Length 349;  
 Best Local Similarity 37.3%; Pred. No. 3.6e-39;  
 Matches 120; Conservative 50; Mismatches 141; Indels 11; Gaps 5;

QY 6 TVAEICGNNGSVDIRRMILKAKKGVNAVKQTRKADKLISALPKAEYQIKNGELE 65  
 DB 22 LIEIQLNHEGSLKTFEWDALVNAKAEVVKQTHIVE---DEMDKQVIRGADV- 77  
 QY 66 SQLEMTKLEMKDYDLHMEYAVSINLDVSTPDESDIDFLASIKOKIWKIPSGELLN 125  
 DB 78 SIYEIMERKALNEDSDIKKEAYESKMTFISTPFRALRLQRMDDIPAYKIGSECCNN 137  
 QY 126 LPLEKIAKLPIDPKKIIITSGATIDEIKOSVITINNKVPVGNITIIHCNTEYTPPE 185  
 DB 138 YPLIKLVASF--GKPIILSTGNSTIESIKSVETIIRKGV--YALHLCNIIYPTPIB 191  
 QY 186 DVNNAINIDLKPKPKNNIGFSDHSGFYAALAAVYGTIFIEKHFTLTKSMGPDHLA 245  
 DB 192 DVRLGMDNLSEKFPDAIIGSDHTLDNVAACGAVALGSLIERHITDRDRGPDIVCS 251  
 QY 246 IEPDELKHLCTIGRCVEKSLGNSKVVLTASERKNKIIVARKSIIAKTEIKKGEVSEKNT 305  
 DB 252 MNPDTFKELKOGHAKLARGKKTIIIGEKPTDPAFASVADDDIKKGLISGDNLM 311  
 QY 306 TKRPGNG-ISPMENYILGKIA 326  
 DB 312 VKRPGNGDPSVNEVETLPGKVA 333

#### RESULT 14

AA41772 ID AAM41772 standard; protein; 400 AA.

AC AAM41772;

DE 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6703.

XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 XX peripheral nervous system; neuropathy; central nervous system; CNS;  
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocastic;  
 XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 XX leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.



Qy 6 IVEIIGCNHNSGDIAIREMIILAKKESGVNAVVEFOFFKADKIISALAPABAOINRTELE 65  
Db 19 VPEIIGINHSDELAKIMVDAFSGAKIIGHOIHIVEDEMSKAKV---PGNAKI- 74  
Qy 66 SOLEMTKIKLEMKVDYLIHMEYAVSLNDIVESTPEDESDIDELASIKOKIKIPSEGLN 125  
Db 75 SIYEIMQKCALDKYDIALKYTEKIKGLVYIISTPSRQANRLBDMGVSAFIKISSGCNN 134  
Qy 126 LPYLEKIAKLPIDPKKIIISTGMATIDEIKOSVSIFFNNKVPVGNITILICNTEYPTPFE 185  
Db 135 YPIKIKHIAF---KKEMIVSTGNNSIESIKPTVKILINDNEIP---FVLMPTTNLYPPHN 188  
Qy 186 DVMINAINDIKKHPPKNNIGSPHSGGFFAALAAVPGITPEKKEFTILDKXSGSDHILAS 245  
Db 189 LVRLINAMLEIKKEP-SCWVGLSDHTINDINALCGAVALACVIERFTDMSHRSGDVLVCS 247  
Qy 246 IEPEDELKHLICIGRVEKSLGSN-SKVVTASERKKIVARRSIIAKTEIKKGEVSEKNI 304  
Db 248 MDIOALKEIIIOSEOMALMRGNNSKKAKOEQVITIDAFASVVISIKDIKKGEVLSYDNI 307  
Qy 305 TTTRPG-NGISPMEWNLGKLAEOFPIDDELIHSEF 341  
Db 308 WTKRPGIGISAAEFENIIGKRLRIENDYOLUSTEDF 345

Search completed: May 6, 2004, 09:04:27  
Job time : 52.7787 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 13.1424 Seconds

(without alignments)  
2532.427 Million cell updates/sec

Title: US-09-930-440b-8

Perfect score: 1778

Sequence: 1 MSNIIYVAEIGCNHNSVDI.....EQDIPDELIHSEFNQGE 346

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1778	100.0	346	2	169836	neub protein - Bsc
2	979.5	55.1	341	2	T44651	capsular polysacch
3	867	48.8	334	2	D81276	N-acetylneuraminic
4	647	36.4	337	2	H64432	spore coat polysac
5	504	28.3	359	2	IC7321	N-acetylneuraminic
6	497.5	28.0	349	2	SC0760	polysialic acid ca
7	493	27.7	350	2	F97169	sialic acid syntha
8	492.5	27.7	343	2	B81275	N-acetylneuraminic
9	472	26.5	373	2	S39722	spore coat polysac
10	443.5	24.9	356	2	D87604	neub protein, prob
11	426	24.0	343	2	H81318	N-acetylneuraminic
12	402	22.6	340	2	B64542	spore coat polysac
13	381	21.4	340	2	B71965	sialic acid syntha
14	266.5	15.0	378	2	H71307	probable spore coa
15	124	7.0	660	2	E90595	hypothetical prote
16	113	6.4	1222	2	B90593	hypothetical prote
17	110.5	6.2	1365	1	BVBYS	killer toxin resis
18	110	6.2	769	2	SS4525	mismatch repair pr
19	109.5	6.2	420	2	B84788	hypothetical prote
20	109	6.1	601	2	C81715	gcpB protein TC032
21	107.5	6.0	713	2	SS6437	probable membrane
22	107.5	6.0	2748	2	S57976	nuclear migration
23	107	6.0	409	2	T43599	yop targeted effec
24	107	6.0	1577	2	T19722	hypothetical prote
25	106.5	6.0	1061	2	C86630	protein F4H10.4 (
26	105.5	5.9	1428	2	SC2419	hypothetical prote
27	105.5	5.9	1620	2	B61535	nucleotide-binding
28	103.5	5.8	623	2	SS6206	probable membrane
29	103	5.8	653	2	A29821	dnax-type molecula

30	103	5.8	732	2	UC6099	Ku autoantigen 66k
31	103	5.8	950	2	E70203	exonuclease Sbc (
32	103	5.8	2269	2	T28677	thoxy protein -
33	102.5	5.8	519	2	E82932	spermidine/putresc
34	102.5	5.8	567	2	F64453	oxalacetate decar
35	102.5	5.8	739	2	JN0581	vascular cell adhe
36	102.5	5.8	1163	2	D64315	type I restriction
37	102.5	5.8	1186	2	C64588	cag pathogenicity
38	102	5.7	1036	2	E96882	hypothetical prote
39	102	5.7	1241	2	F97286	DNA-dependent RNA
40	102	5.7	2136	2	A05037	hypothetical prote
41	101	5.7	857	2	E95009	cell wall surface
42	101	5.7	4688	2	F82885	hypothetical prote
43	100.5	5.7	342	2	H81317	probable lipopolys
44	100.5	5.7	563	2	B82883	hypothetical prote
45	100	5.6	557	2	B83962	hypothetical prote

## ALIGNMENTS

## RESULT 1

169836

neub protein - Escherichia coli

C/Species: Escherichia coli

C/Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 20-Jun-2000

C/Accession: 169836

R:Annunziato, P.W.; Wright, L.F.; Vann, W.F.; Silver, R.P.

J. Bacteriol. 177, 312-319, 1995

A/Title: Nucleotide sequence and genetic analysis of the neub and neub genes in region:

A/Reference number: 155145; MUID:5513767; PMID:7814319

A/Accession: 169836

A/Status: preliminary; translated from GB/EMBL/DD8J

A/Molecule type: DNA

A/Residues: 1346 &lt;RES&gt;

A/Cross-References: EMBL:U05248; NID:9454079; PID:NAC43302.1; PID:9454081

C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpi

Query Match 100.0%; Score 1778; DB 2; Length 346;  
Best Local Similarity 100.0%; Pred. No. 1.7e-117;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSNIIYVAEIGCNHNSVDIAREMILKAKAGVNAVFCFFKADKLISAPRAEYIKN	60
DB	1	MSNIIYVAEIGCNHNSVDIAREMILKAKAGVNAVFCFFKADKLISAPRAEYIKN	60
QY	61	TGELISQLEMTKKLEMYDDYDLHMEYAVSLNDFVSTPPDESDIDFLASKOKIKIPS	120
DB	61	TGELISQLEMTKKLEMYDDYDLHMEYAVSLNDFVSTPPDESDIDFLASKOKIKIPS	120
QY	121	GELLNLPYLAKIAKLPIPDKKIIISTGMATIDBIKOSVSIFINNKPVGNIITILHCTERY	180
DB	121	GELLNLPYLAKIAKLPIPDKKIIISTGMATIDBIKOSVSIFINNKPVGNIITILHCTERY	180
QY	181	PTPEPDVNAIANDLKKHPKXNIGFSDHSGGYAALAAVPGITFIEKFTLDKMSGP	240
DB	181	PTPEPDVNAIANDLKKHPKXNIGFSDHSGGYAALAAVPGITFIEKFTLDKMSGP	240
QY	241	DHLASIEPDELKTLGIVRCVEKSLGNSRVYASERKXIVARKSTIAATEIKKGVSF	300
DB	241	DHLASIEPDELKTLGIVRCVEKSLGNSRVYASERKXIVARKSTIAATEIKKGVSF	300
QY	301	EKKITTRPGNGISPMEWYLLGKIAQDPIPELIIHSEFNQGE	346
DB	301	EKKITTRPGNGISPMEWYLLGKIAQDPIPELIIHSEFNQGE	346

## RESULT 2

T44651

capsular polysaccharide biosynthesis protein cpsN (imported) - Streptococcus agalactiae

C/Species: Streptococcus agalactiae

C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jun-2000

C/Accession: T44651

R:Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; Nitayajarn, A.; Rubens, C.E.  
Submitted to the EMBL Data Library, June 1999

A:Reference number: Z22821  
A:Accession: T44651  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-341 <CHA>  
A:Cross-references: EMBL:AF163833; PIDN:AAD53074.1  
A:Experimental source: strain COH1, serotype III  
C:Genetics:  
A:Gene: cpsN  
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 55.1%; Score 979.5; DB 2; Length 341;  
Best Local Similarity 56.3%; Pred. No. 2e-61;  
Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

4 IYVAFGCHNSVDIAEMILKAEAGNAVKPOTFRADKLISALAPKAYQINNGE 63  
2 VYVIAEGCHNSVDIAEMILKAEAGNAVKPOTFRADKLISALAPKAYQINNGE 61  
64 IESOLETKKLEMKYDYVILMEYAVSLNDVSTFDESDIDFLASLKOKIWKIPSGEL 123  
62 AI QLEMTKRLSEFSEYLEMRDYALSKVEFTSTFDESDIDFLSTOMPDIKIPSGEI 121  
DB 124 LKPYLEKAKLPIPKKIITISGMATIDIKOSVSIPIFNKVPVGNITITLHNTYPTP 183  
DB 122 TMLPYLEKIGK--QOKVILSTGMVMEIHCQVNLKONGTT--DISILCTTEYPTP 176  
DB 184 FEVNLNAINDLKCHFPKNNIGSDSHSGFYAIAVPGITPIEKGFTLDKSMGPDHL 243  
DB 177 YPSLNINVLHTLDCFKDLITIGSDHSISEVPINAAAGAEVIEGHFTLDTMBGPDHK 236  
DB 244 ASIEPELKLHLCIGRCVEKSLGDSNRVVTASERKNTVARKSIITAKTEIKGEVSEKN 303  
DB 237 ASATPPIIALVGVIVGIVVEQALRFEPKIPDVEEKRIYARKSVVALKIKKIDIVSIEN 296  
DB 304 ITTKRPGNGISPMWYNLKGIAEODFIPELITISEPNKO 344  
DB 297 ITVKRPGNGISPMWYNLKGIAEODFIPELITISEPNKO 337

RESULT 3  
DB1276  
N-acetylneuraminic acid synthetase (EC 4.1.3.-) C1327 (imported) - Campylobacter jejuni

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: DB1276  
R:Parikh, J.; Wren, B.W.; Mangall, K.; Kelsey, G.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Randleam, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals h  
A:Reference number: AB1250; NCID:20150912; PMID:10688204  
A:Accession: DB1276  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-334 <PAR>  
A:Cross-references: GB:AL139078; GB:AL111168; NID:96368723; PIDN:CAB73754.1; PID:9636878  
A:Experimental source: serotype O2, strain NCTC 1168  
C:Genetics:  
A:Gene: neuB2; C1327  
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

C:Keywords: Carbon-carbon lyase; oxo-acid-lyase  
Query Match 48.8%; Score 867; DB 2; Length 334;  
Best Local Similarity 51.5%; Pred. No. 1.5e-53;  
Matches 173; Conservative 62; Mismatches 97; Indels 4; Gaps 2;

DB 1 MSNIVVAEAGCHNSVDIAEMILKAEAGNAVKPOTFRADKLISALAPKAYQIN 60  
DB 1 MKKTLIAEGVNHGDIINAKKLIETLAADSGADPVKFSFKAKNCISTKAKKAPYQIKT 60  
DB 61 TGELESQLKEMKYDYVILMEYAVSLNDVSTFDESDIDFLASLKOKIWKIPIS 120

DB 61 TANDESQOMVQKELDKAEKELIHAKKKNIAPLSTPFDISEVDLINEGLKIFKIDS 120

DB 121 GELLNLPYLEKAKLPIPKKIITISGMATIDIKOSVSIPIFNKVPVGNITITLHNTYPTP 180  
DB 121 GELLNLPYLEKAKLPIPKKIITISGMATIDIKOSVSIPIFNKVPVGNITITLHNTYPTP 177  
DB 181 PTFEDVALNAINDLKCHFPKNNIGSDSHSGFYAIAVPGITPIEKGFTLDKSMGPDHL 240  
DB 178 PAFNENVLKAMQSLKQAF-KLDVGYSDHTGIRHISLAVALGACVIEKHFTLDKSMGPDHL 236

DB 241 DHALISZDELKHLCTIGRCVEKSLGDSNRVVTASERKNTVARKSIITAKTEIKGEVSEKN 300  
DB 237 DRAKLEEQELKMLCTQROIKAVGDIKAKSSEQNNINIVKSLVAKDKIKKELFS 296  
DB 301 EKNITTKRPGNGISPMWYNLKGIAEODFIPELITISEPNKO 336  
DB 297 EGNLTTKRPANGISPMWYNLKGIAEODFIPELITISEPNKO 332

RESULT 4  
DB6432  
spore coat polysaccharide biosynthesis protein E homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: H6432  
R:Bull, C.J.; Overbeek, R.; Kikness, E.B.; Weinstock, K.G.; Merrick, J.M.; Glöck, A.  
; Reich, C.J.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese,  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; NCID:96337999; PMID:8688087  
A:Accession: H6432  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-337 <SUB>  
A:Cross-references: GB:U67549; GB:L77117; NID:92826363; PIDN:AAB99068.1; PID:91591717;  
A:Map position: REV1006919-1005906  
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 36.4%; Score 647; DB 2; Length 337;  
Best Local Similarity 42.2%; Pred. No. 4.4e-38;  
Matches 140; Conservative 56; Mismatches 116; Indels 20; Gaps 5;

DB 5 YVIAEGCHNSVDIAEMILKAEAGNAVKPOTFRADKLISALAPKAYQINNGE 64  
DB 18 FIAAGSLNHGDIIDIGELVNAKCCADAKRPSYHTEDFIS--KKSEY----- 66  
DB 65 ESQLEMTKLEMKYDYVILMEYAVSLNDVSTFDESDIDFLASLKOKIWKIPSGEL 124  
DB 67 --YLFPSLSESEFELKRYAKIGIMFISFPLDKYVDILNKNVPAFKIASGDLT 123  
DB 125 NLPLYEAKLPIPKKIITISGMATIDIKOSVSIPIFNKVPVGNITITLHNTYPTP 184  
DB 124 FVPLLEKAKL--TKPVLITSGMSDISEIWEAVVLENN--GCRDITLHNTISYPTP 178  
DB 185 EDVNLNAINDLKCHFPKNNIGSDSHSGFYAIAVPGITPIEKGFTLDKSMGPDHL 244  
DB 179 EDVNLNAINDLKCHFPKNNIGSDSHSGFYAIAVPGITPIEKGFTLDKSMGPDHL 237  
DB 245 STPEPELHLCIGRCVEKSLGDSNRVVTASERKNTVARKSIITAKTEIKGEVSEKN 304  
DB 238 SADPEPEPEEMNNIRLVKRMKLSGKILPSESDVITAEARSIVAKNKIKGEVITSDNI 297  
DB 305 TTKRPGNGISPMWYNLKGIAEODFIPELITISEPNKO 336  
DB 298 SFKRRGRIETKYSIILNKKIKNDKEDDII 329

RESULT 5  
JC7321

N-acetylneuraminic acid 9-phosphate synthase (EC 4.1.3.-) - mouse  
 NAlternate names: N-acetylneuraminic acid synthase homolog  
 C:Species: Mus musculus (house mouse)  
 C:Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 06-Oct-2000  
 C:Accession: J07321  
 R:Nakata, D.; Close, B.E.; Colley, K.J.; Matsuda, T.; Kitajima, K.  
 Biochem. Biophys. Res. Commun. 273, 642-648, 2000  
 A:Title: Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphate  
 A:Reference number: J07321  
 A:Accession: J07321  
 A:Residues: 1-359 <NA>  
 A:Molecule type: mRNA  
 A:Cross-references: DDBJ:AB041263  
 A:Comment: This enzyme is a cytosolic enzyme and as a housekeeping enzyme as well, which  
 catalyzes the conversion of N-acetylneuraminic acid 9-phosphate  
 C:Genetics:  
 A:Gene: stalic acid  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 28.3%; Score 504; DB 2; Length 359;  
 Best Local Similarity 36.0%; Pred. No. 5.3e-28;  
 Matches 124; Conservative 59; Mismatches 145; Indels 16; Gaps 7;

QY 5 YVVAIGCNHNSVDIAEMTLKAKAGVNAVKQTFKADKLISALAPKAEYQIKNTGEL 64  
 DB 20 FIIAEIGNHQGDIDVAKMTIRAKBCGADCAKQKSELBFKRNKALEREYTSKHSWG- 78  
 QY 65 ESQLEMTKLEMKYDDYVLMMEYASVNLDPFSTPDEDSIDFLASLKQKTKWIPSGEL 124  
 DB 79 KYGGHKLHLSHQYELQVLAQELGIFFTAGMDMAVEFHELVNPFKXGSGGTN 138  
 QY 125 NLPLYEKIAKLPIPKKIIISTGNATIDELIKOSVIFINNVPVGNITILHCNTEYPTPE 184  
 DB 139 NPLYLEKAKKGRP---WISSGQSMQMTMGVQIV---KPLNPFELCTATYPLQRP 192  
 QY 185 EDVNLAINDLKQHPKKNIGSPDHSGFYAALAAVPGITFIEKHFTLDKMSGPDHLA 244  
 DB 193 EDANLRVISEYQKFPDPIGYSCHETGIALSVAAVAGATVLRHTLDTMGVSGHSA 252  
 QY 245 SIEPDELKHLICGRVCCKSGNSKVTASERKNTVARKSIIIAKTEIKKGEVSEKNT 304  
 DB 253 SLREELAEVSRVLRALASPKQKLLPCEMACNKGSGSVARKKIPAGTTLTIDTL 312  
 QY 305 TTK--RPGNGISPMEMYNLLAK---IAEDFIPDELI-THSE 340  
 DB 313 TVKVGEP-KGYPPEDIFVLAGKTVLVTIEDDTWAESEVESHK 355

## RESULT 6

60760  
 A:Title: Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid  
 A:Reference number: S60758; MUID:95131727; PMID:7830552  
 A:Accession: S60760  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-149 <ED>  
 A:Cross-references: EMBL:M95053; NID:9520732; PIDD:AA20477.1; PID:9520735  
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, May 1994.  
 R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.B.; Eisen, J.A.  
 Hickey, E.K.; Hatt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 Ri, H.; Olin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massimiani, V.; Piatz, M.  
 Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ve  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain M658.  
 A:Reference number: A51000; MUID:2017535; PMID:10710307  
 A:Accession: B81241

A:Molecule type: DNA  
 A:Residues: 1-349 <IT>  
 A:Cross-references: GB:AE002366; GB:AE002098; NID:97225284; PIDD:AAF40535.1; PID:972252  
 A:Experimental source: serogroup B, strain M658  
 C:Genetics:  
 A:Gene: NMB0068  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cp

Query Match 28.0%; Score 497.5; DB 2; Length 349;  
 Best Local Similarity 37.3%; Pred. No. 1.5e-27;  
 Matches 120; Conservative 50; Mismatches 141; Indels 11; Gaps 5;

QY 6 IVAIGCNHNSVDIAEMTLKAKAGVNAVKQTFKADKLISALAPKAEYQIKNTGEL 65  
 DB 22 IICIGIHNGSLKTAEDVDAVNGAEVYKQTHVE---DEMSDEAKQVIRGADV- 77  
 QY 66 ESQLEMTKLEMKYDDYVLMMEYASVNLDPFSTPDEDSIDFLASLKQKTKWIPSGEL 125  
 DB 78 SIYEIMERKALNEDDEIKLAEVSEKMFISTPFRALALRLQRMIDIPAYKIGSGCEN 137  
 QY 126 LPLYEKIAKLPIPKKIIISTGNATIDELIKOSVIFINNVPVGNITILHCNTEYPTPE 185  
 DB 138 YPLIKLVASF---GRPIISTGNSTISIKSYEILREAGV---YALHCTNLYPTPE 191  
 QY 186 DVNLAINDLKQHPKKNIGSPDHSGFYAALAAVPGITFIEKHFTLDKMSGPDHLA 245  
 DB 192 DVRLGMDNDSEAFPDALIGLSDHTLDNVAALGVALGSGILRHFRDRDRPDPVCS 251  
 QY 246 IEPDELKHLICGRVCCKSGNSKVTASERKNTVARKSIIIAKTEIKKGEVSEKNT 305  
 DB 252 KNPDTFELKQGAHALALAGKQKOTIIAGSEKTKDPAFASVADKDKGSLSGNTW 311  
 QY 306 TKRPGNG-ISPMEYNLLAKIA 326  
 DB 312 VKRQNGDPFVNRVETLFGKVA 333

## RESULT 7

697169  
 A:Title: Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: F97169  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-350 <KR>  
 A:Cross-references: GB:AE001437; PIDD:AAK80145.1; PID:915025183; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC2187  
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 27.7%; Score 493; DB 2; Length 350;  
 Best Local Similarity 35.7%; Pred. No. 3e-27;  
 Matches 117; Conservative 69; Mismatches 130; Indels 12; Gaps 7;

QY 5 YVVAIGCNHNSVDIAEMTLKAKAGVNAVKQTFKADKLISALAPKAEYQIKNTGEL 64  
 DB 19 FIIAEISAHNNDPDAVEITIKAKSGADALITQITTPD-TITPDSDNKYFOIKQ-GTI 76  
 QY 65 ESQLEMTKLEMKYDDYVLMMEYASVNLDPFSTPDEDSIDFLASLKQKTKWIPSGEL 121  
 DB 77 MDCTTHLKHLYEEAVTQWQPKLKEIAEEGGLCFSSPPDNTSVDFLEKNEVAYKVASF 136  
 QY 122 ELNLPLYEKIAKLPIPKKIIISTGNATIDELIKOSVIFINNVPVGNITILHCNTEYPT 181  
 DB 137 ELNDIPFIRIIS---KGRVIVSTGIALNGEITVEAVAAC--KRGGENVITLKCSSSY 191

QY 182 TPEEDVNLAINDLKHPKNNIGFSDHSGFYAAIAYVYGI:FTLEKHTLTKSMSPD 241  
 192 SPEDINLKTIPKMRERAF-NCSVGLSDHTMGYSVALAAVALGATVIEKFTLKRSDDGPD 250  
 QY 242 HIASIPDELKHLCTGRCVCKSLGNSKRVATASERKNIVAKSIIATETIKGVSFSE 301  
 251 SAFSMPEEBSAAVAKSIREVEKLGKVTYELTEKQNSRQHSRLFPVK-DIKKGETFTK 309  
 QY 302 KNITKRPNGISPMEMYNLKGIAEOD 329  
 310 ENVKSIRPAFGTKTYIEVIGKKAAYD 337

RESULT 8  
 B81275  
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1317 [imported] - Campylobacter jejuni  
 C/Species: Campylobacter jejuni  
 C/Date: 31-Mar-2000 #sequence, revision 31-Mar-2000 #ext, change 03-Jun-2002  
 C/Accession: B81275  
 R/Parkhill, J.; Wren, B.W.; Mungall, K.; Kestley, J.M.; Churcher, C.; Basham, D.; Chillingworth, W.; Quail, M.; Randleman, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, N.; et al. The complete genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf  
 A/Reference number: A81250; MUID:20150912; PMID:10688204  
 A/Accession: B81275  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-343 <PAR>  
 A/Cross-references: GB:AL119078; GB:AL111168; NID:G6968723; PIDN:CA873744.1; PID:G696875  
 A/Experimental source: serotype O2, strain NCTC 11166  
 C/Genetics:  
 A/Gene: neuB3, Cj1317  
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 27.7%; Score 492.5; DB 2; Length 343;  
 Best Local Similarity 35.3%; Pred. No. 3.2e-27;  
 Matches 120; Conservative 67; Mismatches 144; Indels 9; Gaps 6;  
 QY 4 IYVAEIGCNHNSVDIAREMILKAEAGVNAVRFQTFKADKLISALAPKAEYQIK-NTG 62  
 12 VFIIAEISANHAGSLKALSKIRAKAGADAIKIQTYPDSE-TINSDEKEDPIIKGILM 70  
 QY 63 ELDSQLEMTKLEMKYDDYLHMEVAVSLNLDVFSFPEDESDIFLASLKQIKWIPGE 122  
 71 DKRKLYELVESATPYEMHSQIFETIIONGILCFSSPMAEDVEFLKRPDPAYKIASFE 130  
 QY 123 LNLPLYLEKIAKLPIPDKKIISTGMATIDEIKOSYSIFENNVAPGNITILHCNTEPT 182  
 131 ANDENFVRLAK--EKPTIVSTGATSEELFKICEIFEEKNP--DVFLEKSTSTYPT 185  
 QY 183 PFEDVNLAINDLKHPKNNIGFSDHSGFYAAIAYVYGI:FTLEKHTLTKSMSPD 242  
 186 AIEDMMKLGIVSKERF-NVEVGLSDHSFGFLAPVAVAGARVYKHEHMLDKSLSESES 244  
 QY 243 LASISEDELKHLCTGRCVCKSLGNSKRVATASERKNIVAKSIIATETIKGVSFSE 302  
 245 KPSLDIDPEFAMVDARQASALAGDGKLDDEKVLKNRYPAR-SLYASNDIKKGEMFSSE 303  
 QY 303 NITTKRPNGISPMEMYNLKGIAEODPIPEDELIHSEK 342  
 304 NVKSVSPFGLHFKFYQELLGKKAAYDIFKGDALCKGDFQ 343

RESULT 9  
 S39722  
 spore coat polysaccharide synthetase protein spse - Bacillus subtilis  
 N/Alternate names: protein lpa-66d  
 C/Species: Bacillus subtilis  
 C/Date: 07-Oct-1994 #sequence, revision 26-May-1995 #ext, change 20-Jun-2000  
 C/Accession: S39722; C69717  
 R/Glaser, P.; Kunst, F.; Arnaut, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,

A.; Rapoport, G.; Danchin, A.  
 Mol. Microbiol. 10, 371-384, 1993  
 A/Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region f  
 A/Reference number: S39655; MUID:95020537; PMID:7534828  
 A/Accession: S39722  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-373 <GLA>

A/Cross-references: EMBL:X73124; NID:G413923; PIDN:CA451623.1; PID:G413991  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993  
 R/Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
 C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Ch  
 A.; Enrich, S.D.; Emerson, P.T.; Enliat, K.D.; Errington, J.; Fabret, C.; Ferrati, E.  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Galle  
 jech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinois  
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y., M.; Ogawa, K.; Ogatawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Poterell  
 Rieger, W.; Rivolta, A.; Rocha, E.; Roche, B.; Rose, M.; Sadde, Y.; Sato, T.; Scanlon  
 A/Authors: Schleich, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Senc  
 akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Toso, V.; Uchiyama  
 T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,  
 A/Authors: Yoshikawa, H.F.; Zumbstein, B.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: C69717  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-373 <KUN>  
 A/Cross-references: GB:Z99123; GB:AL009126; NID:G2636240; PIDN:CA815813.1; PID:G2636324  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Gene: spsB  
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein ci

Query Match 26.5%; Score 472; DB 2; Length 373;  
 Best Local Similarity 36.4%; Pred. No. 9.8e-26;  
 Matches 132; Conservative 55; Mismatches 128; Indels 48; Gaps 11;  
 QY 4 IYVAEIGCNHNSVDIAREMILKAEAGVNAVRFQTFKADKLISALAPKAEYQIKNTGE 63  
 17 VFIIAEISANHAGSLKALSKIRAKAGADAVFQWFOQAQRM-----YQ-KDQGL 66  
 QY 64 LE-----SQLEMTKLEMKYDDYLHMEVAVSLNLDVFSFPEDESDIFLASLKQIKW 116  
 67 YKTAAGQDVSTFSLQSNEMPAWILPLDTCREKQVIFLSTVCDEGADLLQSTSPAF 126  
 QY 117 KIPSGELNPLYLEKIAKLPIPDKKIISTGMATIDEIKO--SVSIFINKVPGNITI 173  
 127 KIASYEINHLPLKIVARL--NRPMISTGAEISDVHEAMRTIRAGNNO-----TAI 178  
 QY 174 LHCNTEPTPEEDVNLAINDLKHPKNNIGFSDHSGFYAAIAYVYGI:FTLEKHTLTKSMSPD 242  
 179 MECAVAPAPPEYSLSYSTPLAAAPFAVAVGFSHSEHPLEAPCAARLQKLIKFT 238  
 QY 233 LDKSMSPGDHLASIEPDELKHLCTGVR-----CVKSLGNSKRVATASERKN 279  
 239 IDKUNPAGDHFPALNPDLDKEMVNGIRTBALQKIGITKAPSEKLLGSSYKTTVAIBEEL 298  
 QY 280 KIVAKSIIATETIKGVSERKNITTKRG--NGISPMEMYNL--GKIAEODFIDE 334  
 289 RNFARGLFTTAIFQKGAESFEDNTAVLRPQKQGLHP-RFELLISGVAVRVIDPDT 357

RESULT 10  
 D87604  
 new protein, probable [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C/Accession: D87604  
 R:Nerman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Leub, M.C.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: D87604  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-356 <STO>  
 A/Cross-references: GB:AE005673; MID:913424484; PIDN:AAK24832.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC2868  
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 24.9%; Score 443.5; DB 2; Length 356;  
 Best Local Similarity 32.5%; Pred. No. 9.2e-24;  
 Matches 109; Conservative 68; Mismatches 141; Indels 17; Gaps 8;  
 QY 5 YIVAEICGNHNSGVDIAREMILKAKAGVNAVKFQTEKADLISALAPKAEYQIKNTGEL 64  
 26 YVCELSGNHNSGIERCLAWDAADTCDAIKQTITAD-TITLDVDRPEKIH--GSL 82  
 QY 65 --ESQLEMTKLEMTKYDYILHMEYAVSLNDVSTPDESDIDFLASKOKIKWIPSG 121  
 DB 83 WDGTLYELVEEATPEPMHAIPEERARQGVTFSSPDETAVDLDSGAEFAKIASF 142  
 QY 122 ELMLPLVIEKIAKLPIDDKIIISTGMATIDEIKOSVIFINMKVPVGNITLHCNTEYP 181  
 DB 143 EAVDLPPIKIAA--AKGKELISTGMALTEKWTALDPAISGAP--GVLLHCVSSTP 197  
 QY 182 TPEEDVNLAINDIKKH--PKNNIGFSDHSGFYAAIAAVPYGITEKHFILDKSMSC 239  
 DB 198 AFDADVAVRIVPDMAAFGCP--IGLSDFPTGTAASVAAVSLGACAVEKHFILRADGG 254  
 QY 240 PDHLASIEPEDELKICGRCVEKSLGNSKVYVASRKKIVAKRKIIKIKGEVF 299  
 DB 255 PDAFSLPEAEFKALVDOTKAAALPRAHYDVGSB-ATSLFRSLVYTPADVAKGEPL 313  
 QY 300 SEKNITTPKPGNGISPMEMYNLLGKIAEQDPIPE 334  
 DB 314 TRAVRSVRPGNGLPADLDKVLAKGATRLDARGE 348

RESULT 11  
 H81318  
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1141 [imported] - Campylobacter jejuni  
 C/Species: Campylobacter jejuni  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C/Accession: H81318  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, Nature 403, 665-668, 2000  
 A/Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf  
 A/Reference number: A81250; MUID:20150912; PMID:10688204  
 A/Accession: H81318  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-343 <PAR>  
 A/Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CAH73396.1; PID:9696857  
 A/Experimental source: serotype O2, strain NCTC 11168  
 C/Genetics:  
 A/Gene: neuB1, Cj1141  
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps  
 C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 24.0%; Score 426; DB 2; Length 343;  
 Best Local Similarity 35.0%; Pred. No. 1.5e-22;  
 Matches 115; Conservative 57; Mismatches 143; Indels 14; Gaps 7;  
 QY 3 NIVYVAELGCHNNSVDIAREMILKAKAGVNAVKFQTEKADLISALAPKAEYQIKNTG 62

DB 14 NPLIIEIGINNHNSGLEIAKLVADAKRAGAKIKQTIVE---DEMSEAGAVIPGNA 70  
 QY 63 ELESQLEMTKLEMTKYDYILHMEYAVSLNDVSTPDESDIDFLASKOKIKWIPSGE 122  
 DB 71 NI-SIEIMEQALNFKDELAKVEYKQGLVYLSFSPRAANRLEDGVSAYKGSGB 129  
 QY 123 LNLPLVIEKIAKLPIDDKIIISTGMATIDEIKOSVIFINMKVPVGNITLHCNTEYP 182  
 DB 130 CNVYPIKIIAQ--KKMIISTGMNLSIEIKPTKILRDYIF--FVLLHTTLLYPT 183  
 QY 183 PREDVNLAINDIKKHPPKNNI-GFSDHSGFYAAIAAVPYGITEKHFILDKSMSCPD 241  
 DB 184 PSHLVRLQAMLELYKEF--NCLYGLSDHTTNLACIGALAGASVLERHPTDMRKGPD 241  
 QY 242 IASIEPDELKICGRCVEKSLGNSKVYVASRKKIVAKRKIIKIKGEVFSE 301  
 DB 242 IVCSHDSSTKDLINQTOEMVLIRGDNKPKLEQVITDPFAASVYSKIDIKGELIM 301  
 QY 302 KNITTRPG-NGISPMEMYNLLGKIAEQD 329  
 DB 302 DNIVVVRPSKGISANDFEALIGKRAKD 330

RESULT 12  
 B64542  
 spore coat polysaccharide biosynthesis protein B - *Helicobacter pylori* (strain 26695)  
 C/Species: *Helicobacter pylori*  
 C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 20-Jun-2000  
 C/Accession: B64542  
 R:Tombs, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khajik, H.G.; Glodok, A.; McKern, son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L. Nature 388, 539-547, 1997  
 A/Authors: Wallin, B.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C. A/Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: B64542  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-340 <TOM>  
 A/Cross-references: GB:AE005328; GB:AE005111; NID:92313263; PIDN:AAD07248.1; PID:9231326  
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 22.6%; Score 402; DB 2; Length 340;  
 Best Local Similarity 31.8%; Pred. No. 7.1e-21;  
 Matches 104; Conservative 74; Mismatches 135; Indels 14; Gaps 8;

QY 6 IYAEICGNHNSGVDIAREMILKAKAGVNAVKFQTEKADLISALAPKAEYQIKNTG-GBL 64  
 DB 7 IYAEICGNHNSGVDIAREMILKAKAGVNAVKFQTEKADLISALAPKAEYQIKNTG-GBL 65  
 QY 65 ESQLEMTKLEMTKYDYILHMEYAVSLNDVSTPDESDIDFLASKOKIKWIPSGEL 124  
 DB 66 ENLYEYORASTPLEVHAELEFELARLDIGIFSSPSQALTELSINCEMYKIASFEIV 125  
 QY 125 NLPLYEKIAPLPIDDKIIISTGMATIDEIKOSVIF--INNKVPVGNITLHCNTEYP 182  
 DB 126 DLDILKAAK--TQPTIILSSGIAHTLEQDASLCRRNN---FDITLKCVAAYS 178  
 QY 183 PREDVNLAINDIKKHPPKNNIGFSDHSGFYAAIAAVPYGITEKHFILDKSMSCPD 242  
 DB 179 KIDANMLLSKVKGELIFGV-KFGLSDHTIGSLPIIATTLGASMIKHPILNKSLOTPDS 237  
 QY 243 LASIEPDELKICGRCVEKSLGNS-KVYVASRKKIVAKRKIIKIKGEVFSE 301  
 DB 238 AFDADVAVRIVPDMAAFGCP--IGLSDFPTGTAASVAAVSLGACAVEKHFILRADGG 254  
 QY 302 KNITTRPGNGISPMEMYNLLGKIAEQ 328  
 DB 297 NNKALRPMLGLHPKRYEITLGGKAK 323



## RESULT 13

B71965 static acid synthase - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Jun-2000

C:Accession: B71965

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557; PMID:99231682

A:Accession: B71965

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-340 &lt;ARN&gt;

A:Cross-references: GB:AE001455; GB:AE001439; NID:G4154678; PIDN:AA05747.1; PID:G415468

A:Experimental source: strain J99

C:Genetics:

A:Gene: neuB

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 21.4%; Score 381; DB 2; Length 340;

Best Local Similarity 30.1%; Pred. No. 2.1e-19;

Matches 98; Conservative 75; Mismatches 141; Indels 12; Gaps 6;

6 IVAEICGNHGSVDIAREMILKAKAGVNAVKEOTFRKADLISALPKAEYOIKYT-GSL 64

7 IVAEISANHQDINLAESLHAIKESGADFKQY-TRSCYLDSEKDEPFIIQGTLMX 65

65 ESOLERTKLEMYDYDYLHMEYAVSLNLEVFSTPEDESIDFLASKQIMKIPSGEL 124

66 ENYGYQKASTPLENHAELFELAKLIDGIFSSPSSKALTELESDCMYKIAFEIV 125

125 NLYLEKIAKLPDPDKIIISTGMATIDEIKOSVSIF--INNKVPVGNITILHNTBYPT 182

126 DLEIEIAAR--TQPRHISGIAITAELODASLCRGVNN---FDITLKCVAAYS 178

183 PEDVAINALINDIKKFPKNNIGFSDHSGFYAIAAVPGITFIEKHFTLDSKSGPDH 242

179 KIDPAHLISWYLGATPGV-KFGSDHTIGSLCPILATLIGASMIKHLNKS-LTPDS 237

243 LASIEPELKHLCIGVRCVEKISGNSKVTYASERKKIVAKSIITAKETIKGEVFSK 302

238 AFEMDNGFMSWGAIKQSVLAGEEPKINPTLIERRPFARSLFVIDIQGEALTS 297

303 NITTKPGNGISPMEMYNILAKIAQ 328

298 NIKALRNILGALHFKYKEIIGQAKS 323

RESULT 14

H71307

Probable spore coat polysaccharide biosynthesis protein (spae) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 19-Apr-2002

C:Accession: H71307

R:Faaser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: H71307

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-378 &lt;COL&gt;

A:Cross-references: GB:AE001232; GB:AE000520; NID:G3322856; PIDN:AA065539.1; PID:G332285

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0562

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 7.0%; Score 124; DB 2; Length 660;

Best Local Similarity 21.8%; Pred. No. 0.59;

Matches 77; Conservative 52; Mismatches 138; Indels 86; Gaps 13;

4 IYIVAEICGNHGSVDIAREMILK-----AKAGVNAVKEOTFRKADLISALPKAEYOIKYT 54

298 VPIPELDFTNN-----TRILLKNNLEKQVNNQNPQNKLSFSSOYELTGFAF-- 350

55 EYQIKNTG--ELISOLETKYLENK--YDYTLHMEYAVSLNLEVFSTPEDESIDFLAS 110

351---INKBEVLEBOELSKIKLSKSKAKNNKLPSQISSTSLKDFQWQSVNGIDFSIE 407

111 LKQIKWKLPSGGLNLPYLEKIAKLPDPDKIIISTGMATIDEIKOSVSIFINNKVPVGNITILH 175

408 LKQGANNNAGLSLIIYARK--KSVESKITTIGFLKQDYDLITINALDSLKNNY 465

171 ITILHNTYPTPEFVNLNAINDLK--HPPKNNIGFSDHSGFYAIAAVPGITFIE 228

466 STLLPSNVE-----INDLEKIVDLPENDGF-----VYGLLEK 499

229 KPEFLDKMSGPDHLASIEPELKHLCIGVRCVEKISGNSKVTYASERKKIVAKSII 288

500 SRFSVN-----DEKGTIK-----VTLQSLKNNLTVEKKDI 529

Query Match 15.0%; Score 266.5; DB 2; Length 378;

Best Local Similarity 26.1%; Pred. No. 2.7e-11;

Matches 90; Conservative 60; Mismatches 152; Indels 43; Gaps 8;

2 SNIVYVAGCNHGSVDIAREMILKAKAGVNAVKEOTFRKADLISALPKAEYOIKYT 61

13 ADILITAGSHAGSFPBRALIDAAADAAAANKPDLITVAHEILHL-----T 62

62 GELE-----SOLETKYLEMYDYDYLHMEYAVSLNLEVFSTPEDESIDFLASKQI 115

63 GAVRLPSGAVSLYQRFEELEVLPSFAOCFNHARSGVLVIGSPGPRSALEALKDF 122

116 KPIPSGELNLPYLEKIAKLPDPDKIIISTGMATIDEIKOSVSIFINNKVPVGNITILH 175

123 LKVASPE-LNTPFL--ISTLAALPILSSGVCILKEITBGLACRCRYTQSGSHALH 179

176 CNTEYPTPEFVNLNAINDLKHPKNNIGFSDHSGFYAIAAVPGITFIEKHFTLD 234

180 CITAYPADTEYNALALPALATIF--NINWGVSDHSDVDLVPFLARAGACIVEKHICLS 238

235 KMSGPDHLASIEPELKHLCIGVRCVEKIS-----IGSNSKYV 272

239 KPDAGLSDSIALDPDFTPTMALNSCARSPSQIISFLHKGVAHPVAVIGSGEVL 298

273 TASEKNNIVAKSIIATYETIKGEVFSKNTTKRPNNGISPYE 317

299 ASESRAHQKSNRSLHYLAAYPGTVLQKENVLYRSEANLSAGE 343

RESULT 15

E90595

Hypothetical protein MYPV 6690 (imported) - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C:Accession: E90595

R:Chhabra, I.; Hallig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul

A:Reference number: A9512; MUID:21267165; PMID:11353084

A:Accession: E90595

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-660 <KUS>

A:Cross-references: GB:AL445566; PID:G14050084; PIDN:CA031482.1; GSPDB:GN00153

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYPV 6690

A:Genetic code: SGC3

Query Match 7.0%; Score 124; DB 2; Length 660;

Best Local Similarity 21.8%; Pred. No. 0.59;

Matches 77; Conservative 52; Mismatches 138; Indels 86; Gaps 13;

4 IYIVAEICGNHGSVDIAREMILK-----AKAGVNAVKEOTFRKADLISALPKAEYOIKYT 54

298 VPIPELDFTNN-----TRILLKNNLEKQVNNQNPQNKLSFSSOYELTGFAF-- 350

55 EYQIKNTG--ELISOLETKYLENK--YDYTLHMEYAVSLNLEVFSTPEDESIDFLAS 110

351---INKBEVLEBOELSKIKLSKSKAKNNKLPSQISSTSLKDFQWQSVNGIDFSIE 407

111 LKQIKWKLPSGGLNLPYLEKIAKLPDPDKIIISTGMATIDEIKOSVSIFINNKVPVGNITILH 175

408 LKQGANNNAGLSLIIYARK--KSVESKITTIGFLKQDYDLITINALDSLKNNY 465

171 ITILHNTYPTPEFVNLNAINDLK--HPPKNNIGFSDHSGFYAIAAVPGITFIE 228

466 STLLPSNVE-----INDLEKIVDLPENDGF-----VYGLLEK 499

229 KPEFLDKMSGPDHLASIEPELKHLCIGVRCVEKISGNSKVTYASERKKIVAKSII 288

500 SRFSVN-----DEKGTIK-----VTLQSLKNNLTVEKKDI 529

Thu May 13 11:53:14 2004

us-09-930-440b-8.rpr

Page 7

QY 289 AKTEIKKGEVSEKNITTKRPGNGISPMENYUJLGIAPEDPFIPELIIHSEF 341  
D5 530 QISGFRITDEFIRONFSSL--GRAIRTRKEGFDLKNYSA-SDLIPREDIPASAY 579

Search completed: May 6, 2004, 09:09:48  
Job time : 16.3924 secs

Thu May 13 11:53:14 2004

us-09-930-440b-8.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:56:04 ; Search time 9.09861 Seconds  
(without alignments)  
1980.112 Million cell updates/sec

Title: US-09-930-440b-8

Perfect score: 1778  
Sequence: 1 MSNIVVAEIGCHNNGSVDI.....EQDFIDELIHSEFNQGE 346

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	647	36.4	337	1	YAE5_METUA
2	516.5	29.0	359	1	SIAS_HUMAN
3	472	26.5	373	1	SPSE_BACSU
4	110.5	6.2	1365	1	KRES_YEAST
5	110	6.2	769	1	MLH1_YEAST
6	109	6.1	601	1	ISPG_CHAMU
7	107.5	6.0	713	1	Y338_YEAST
8	107.5	6.0	2748	1	NUM1_YEAST
9	107	6.0	409	1	YOPM_YEPE
10	105.5	5.9	1428	1	YAB4_SCHPO
11	103.5	5.8	623	1	YEP3_YEAST
12	102.5	5.8	567	1	PYC2_YEAST
13	102.5	5.8	739	1	YCAL_MOUSE
14	102.5	5.8	1075	1	Y124_METUA
15	102.5	5.8	1186	1	CAGA_HELPY
16	102	5.7	1241	1	RPOB_CLOAB
17	102	5.7	2136	1	YCF2_MARPO
18	100	5.6	878	1	MSH4_YEAST
19	99.5	5.6	887	1	TOP1_BACAN
20	98.5	5.5	373	1	ASPM_MOUSE
21	98.5	5.5	654	1	PGKT_THEMA
22	97.5	5.5	332	1	NDPA_VIBPA
23	97.5	5.5	962	1	IFJ3_MAIZE
24	97.5	5.5	969	1	DP3A_URAPA
25	97.5	5.5	1460	1	NIS9_YEAST
26	97.5	5.5	5171	1	BPEA_HUMAN
27	97	5.5	554	1	Y478_RICPR
28	96.5	5.4	409	1	Y4DM_RHLSN
29	96.5	5.4	1138	1	CTAA_BACTU
30	96.5	5.4	1782	1	VIT_BOMMO
31	96.5	5.4	1862	1	GSAB_PICPA
32	96	5.4	426	1	PUR2_FUSNM
33	96	5.4	603	1	UVRC_LISMO

34	96	5.4	1169	1	SMC_METUA	059037 methanococ
35	95	5.3	375	1	FLG1_BUCBP	089485 buchnera ap
36	94.5	5.3	597	1	NFL2_RAT	054968 ractus norv
37	94.5	5.3	646	1	SR72_YEAST	038688 saccharomyc
38	94.5	5.3	2867	1	RBP2_PLAVB	000799 plasmodium
39	94	5.3	336	1	Y625_METUA	058042 methanococ
40	94	5.3	395	1	PUR2_STNP7	054978 synecococc
41	94	5.3	501	1	MTB8_BACSU	033563 bacillus su
42	94	5.3	798	1	SYR2_BUCAP	059057 buchnera ap
43	94	5.3	1583	1	GCC2_HUMAN	081WJ2 homo sapien
44	93.5	5.3	415	1	Y310_BUCAP	08K9M4 buchnera ap
45	93.5	5.3	729	1	KAR3_YEAST	017119 saccharomyc

## ALIGNMENTS

### RESULT 1

ID	YAE5_METUA	STANDARD	PRT	337 AA.
AC	Q58465			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Hypothetical protein MJ1065.			
GN	MJ1065.			
OS	Methanococcus jannaschii.			
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;			
OC	Methanocaldococcaceae; Methanocaldococcus.			
OX	NCBI_TaxId=2190;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;			
RX	MEDLINE=96317999; PubMed=868087;			
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,			
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,			
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,			
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,			
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.D., Sadow P.W., Hanna M.C.,			
RA	Uttreback I.R., Kelley U.M., Peterson J.D., Sadow P.W., Bordovsky M.,			
RA	Colton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,			
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;			
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus			
RT	jannaschii.";			
RL	Science 273:1058-1073(1996).			
CC	- SIMILARITY: STRONG, TO B. SUBTILIS SPSE.			
CC	- SIMILARITY: Contains 1 AFP-like domain.			
CC	- SIMILARITY: Contains 1 AFP-like domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL: U67549; AAB9068.1; -			
DR	PIR: H64432; H64432.			
DR	TIGR: MJ1065;			
DR	InterPro: IPR006014; Antifreeze dom.			
DR	InterPro: IPR006190; Antifreeze-like.			
DR	InterPro: IPR004144; Neut.			
DR	Pfam: PF01354; Antifreeze; 1.			
DR	Pfam: PF03102; Neut; 1.			
DR	PROSITE: PS50844; AFP_LIKE; 1.			
KW	Hypothetical protein; Complete proteome.			
FT	DOMAIN 279 337 AFP-LIKE.			
SQ	SEQUENCE 337 AA; 37976 MW; 790CCAF46C111B CRC64;			
Query Match	36.4%; Score 647; DB 1; Length 337;			
Best local similarity	42.2%; Pred. No. 3.7e-37;			
Matches 140; Conservative 56; Mismatches 116; Indels 20; Gaps 5;				

RESULT 2	
SIAS_HUMAN	
ID_SIAS_HUMAN	STANDARD; PRT; 359 AA

RN	[2]
RP	SEQUENCE FROM N.A.

RC	TISSUE=lung, and Placenta;
RX	MEDLINE=22388257; PubMed=12477932;
PA	Stromberg B I, Fainberg A, Cr

RA Slatkowski R.L., Jastkowski P.A., Grobe L.H., DeGee G.C.  
RA Klausner R.D., Collins F.S., Wagner L., Siemien C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Boretow R.H., Schaefer C.F., Hsieh F.,  
RA Hopkins R.F., Jordan H., Moore T.K., Sax S.I., Wang J., Hatt N.K.,  
RA Ditchenko L., Maruskin K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stepleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Iannelli N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosnak S.A., McBean P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Harte S., Garcia A.M., Gay L.J., Hultyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```
CC -----
DR EMBL, AF257466, AAF5261.1, -
DR EMBL, AK001659, BAA9181.1, -
DR EMBL, BC000008, AA000008.1, -
DR EMBL, BC019315, AAA19315.1, -
DR MIM, 605202, -
DR HSSB, P19614, NAME.
DR GO, GO:0005737, C:cytoplasm, NAS.
```

DR	GO:0008781; F: N-Acetylserine aminotransferase acti. . .; NAs.
DR	GO:0009103; P:11ppolysaccharide biosynthesis; NAs.
DR	InterPro; IPR006014; Antifreeze_dom.
DR	InterPro; IPR006190; Antifreeze_like.
DR	InterPro; IPR004144; Neub.
DR	Pfam; PF01354; Antifreeze; 1.
DR	Pfam; PF03102; Neub; 1.
DR	ProDom; PD003258; AntifreezeII; 1.
DR	ProSite; PS50844; AFP_LIKE; 1.
KW	Transferase; Polymorphism.
FT	DOMAIN 294 353 AFP-LIKE.
FT	VARIANT 68 68 E->D.
FT	FTID=VAR_013308.
FT	CONFLICT 232 232 A->T (IN REF. 2).
FT	CONFLICT 321 321 G->A (IN REF. 1).
SC	SEQUENCE 359 AA; 40307 MM; 2E02D547F4F96592C CMC64;

Query Match	29.0%	Score 516.5	DB 1	Length 359
Best Local Similarity	36.6%	Pred. No. 3.2e-28		
Matches 124	Conservative 61	Mismatches 139	Indels 15	Gaps 6
QY	5	YTVAEIGNHNGSVDIAREMTLLAKKAEGVNVAVKFQTEPAADKLISAIAPKAEYQIKNTGEL	64	
DB	20	FLIETIGNHGGDLDVAAGMTIRMAKCEACADCAKQKSLERKFKALERTYSTSHSG-	78	
QY	65	ESQLEMTKKELMKRYDDYLHLMEEYVAIVNLDFVSPDEDSIDFTASLKKQIKWJPSGELL	124	
DB	79	KTYGHHKHLTFESHQYRELQRYAAEYVGIPTLGAQMDMAVEFLHMLVAFKVGSSGDTN	138	
QY	125	NLPYIEKIAKPIPKKIIITGMAITDEICQSSSIFLNKNVAPYGNIIILCNTEYPPPF	184	
DB	139	NFPYIEKIAKKGPR--WVSSSQMSQMTKMQVQIV--KPLANPFCFLQCTSAIVPQOP	192	
QY	185	EDVNLINAIINDLKKIIFPKNNKIGFSDHSSGSGFYAIAAVPYGITEFKHPTLIDKSMGSDHILA	244	
DB	193	EDVNLRYVISEYQKLPFDIPIGYSGHETISALSVAAVAALGAKVLESHITLIDKTMKQSDHSA	252	

Oy	245	SIPELHELICIGVACVCKSKGNSKNKWTASERKKIVARKKIIINTEIKKEVSEKKI	300
Db	253	SBERELAEIVRSVRLVERALOSPFKROLPCMACNCKSGSVAKYKIPETILTMDLT	312
Oy	305	TTK--RPENGISPEWVYNLLGK----IAEQDFIDELI	336
Db	313	TVKVGEP-KGYPPEDIFNVLGKKVLWVEEDTIMEELV	350

RESULT 3

SPSE	BACSU	STANDARD;	PRT;	373	AA.
ID	SPSE	BACSU			
AC	P39625;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Spore coat polysaccharide biosynthesis protein spsz.				
GN	SPSZ OR IPA-67D OR BSU37870.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxId:1423;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RC	MEDLINE=95020537; PubMed=7934828;				
RX	Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,				
RA	Hullo M.F., Ionescu M., Imbouchinsky B., Marcelino L., Mosser I.,				
RA	Presseau E., Santana M., Schneider E., Schweizer U., Vertes A.,				
RA	Rapoport G., Danchin A.;				
RT	"Bacillus subtilis genome project: cloning and sequencing of the 97				
RT	kb region from 325 degrees to 533 degrees."				
RL	Mol. Microbiol. 10:371-384(1993).				
RP	(2)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RC	MEDLINE=968044033; PubMed=9384377;				
RX	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Berttero M.G., Bessieres P., Bolotin A., Borchert S.,				
RA	Borries R., Boutsier L., Brans A., Braun M., Brignelli S.C., Bron S.,				
RA	Brouillet S., Burscht C.V., Caldwell B., Capuano V., Carter N.K.,				
RA	Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,				
RA	Denizot F., Deville K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,				
RA	Enrian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,				
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,				
RA	Guisept G., Guy B.J., Haga K., Hahsch U., Harwood C.R., Henaut A.,				
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,				
RA	Joris B., Karmata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koeltter P., Koningsreim G., Krogh S., Kumano M.,				
RA	Kurita K., Lapidas A., Lardinois S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Mabel C., Medigue C.,				
RA	Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,				
RA	Noone D., O'Reilly W., Ogawa K., Ogiwara A., Oudega B., Park S.H.,				
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,				
RA	Presseau E., Puig P., Purnelle B., Rapoport G., Ray M., Reynolds S.,				
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,				
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,				
RA	Seitguth U., Sekowska A., Seter S.U., Serrir P., Shin B.S., Soldo B.,				
RA	Sorokin A., Taccoti B., Takagi T., Takahashi H., Takemaru K.,				
RA	Takenuchi M., Tamakoshi A., Tanaka T., Terpeira P., Vassarotti A.,				
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,				
RA	Viari A., Wambut R., Wedler S., Wedler H., Westmegeger T.,				
RA	Winters P., Wipar A., Yamamoto H., Yamane K., Yamamoto K., Yata K.,				
RA	Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,				
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus				
RT	subtilis."				
RL	Nature 390:249-256(1997).				
CC	-1- PATHWAY: Spore coat polysaccharide biosynthesis.				
CC	-1- SIMILARITY: Contains 1 AFP-like domain.				

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CAB51623.1; -
DR EMBL; Z99123; CAB1813.1; -
DR PIR; S39722; S39722.
DR Subtilist; BG10613; spsE.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR004144; Neut.
DR Pfam; PF01354; Antifreeze_1.
DR Pfam; PF03102; Neut; 1.
DR PROSITE; PS50844; AFP_Like; 1.
KW Complete proteome.
KW DOMAIN
FT 305 AFP-LIKE.
SQ SEQUENCE 373 AA; 40889 MW; 3CFBDFACCE0DBCB8 CRC64;
Query Match 26.5%; Score 472; DB 1; Length 373;
Best Local Similarity 36.4%; Pred. No. 3.6e-25;
Matches 132; Conservative 55; Mismatches 128; Indels 48; Gaps 11;
QY 4 IYVAIEGCHNGSVDIAREMILKAKAGVNAVYQFFKADKLISALAPKAYGIKNTGE 63
Db 17 VFIIAEGINHDDGLDPAFALIDAAAGADAVYQFQPADRM-----YQ-KDGL 66
QY 64 LE-----SQLEMTKLEMKVDYDLHMEVAVSLINDVSTPDEDSDIFLASLKQIM 116
Db 67 YKTAAGDVGILFSIVSGMEVPAEKLILPLDYCEKQVILPSTVDEBSADILQSTPSAF 126
QY 117 KIPSGELLNPYLKIAKDPIDKKIILISGMATIDEIKQ---SVSIFINNKVPGVNTI 173
Db 127 KIASYEINHLPLKLYVARL---NRPMIFSTAGAEISDVHEAMRTIRAEGNQ---IAI 178
QY 174 LHCNTEVPTPEPDNINAINIDLKKHPKNNIGFSDBSS-GFYAALAAVPYGITFEKHT 232
Db 179 MGCYAKFAPPEPVSNLSVIMPLAALAPPAEAVIGFSDBSHETPAFCAAVRLGAKLIEHFT 238
QY 233 LDNSMSGPDHLASIEPDELRLHLGIVR-----CVKSLGSSNKVVTASERKN 279
Db 239 IDKKLPGADSPFALNPDELKEMVDGIRKTEALIKQGITKPVSEKLGLSSYKTTAIGRI 298
QY 280 KIVAKSKIIAKTELKGEVSESKITTKRG---NGISPMHWYLL--GSLAQDFFIDE 334
Db 299 RNPFYRGIIPTAPIQGEAFSEENIAVLRQGXKPKQGLHP-RFEELLTSGVPAADIPADT 357
QY 335 LII 337
Db 358 GIV 360
RESULT 4
KRES_YEAST STANDARD; PRT; 1365 AA.
AC P23023; Q12190;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Killer toxin-resistance protein 5 precursor.
KRES OR YOR336M.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_Acid=4932;
RX MEDLINE=90258892; Pubmed=2188106;
RA Meaden P., Hill K., Wagner J., Stipetz D., Sommer S.S., Bussey H.;
RT "The yeast KRES gene encodes a probable endoplasmic reticulum protein
required for (1--5)-beta-D-glucan synthesis and normal cell growth."
WL Mol. Cell. Biol. 10:3013-3019(1990).
[2]

```



DR InterPro: IPR003594; Atgbind\_ATPase.  
 DR InterPro: IPR002099; DNA\_mis\_repair.  
 DR Pfam: PF01119; DNA\_mis\_repair; 1.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR SMART: SM00387; HATPase\_C; 1.  
 DR TIGRPFAMs: TIGR00585; nuc1; 1.  
 DR PROSITE: PS00058; DNA\_MISMATCH\_REPAIR\_1; 1.  
 DR DNA\_repair; Nuclear protein.  
 FT MUTAGEN 41 41 A->F: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 41 41 A->S: FULLY FUNCTIONAL IN A MISMATCH  
 FT MUTAGEN 64 64 G->R: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 65 65 I->N: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 99 99 E->K: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 104 104 I->R: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 114 114 T->R: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 214 214 R->C: PARTIALLY DEFECTIVE IN A MISMATCH  
 FT MUTAGEN 216 216 V->I: FULLY FUNCTIONAL IN A MISMATCH  
 FT MUTAGEN 265 265 R->C: PARTIALLY DEFECTIVE IN A MISMATCH  
 FT MUTAGEN 265 265 R->H: PARTIALLY DEFECTIVE IN A MISMATCH  
 FT MUTAGEN 265 265 R->A: PARTIALLY DEFECTIVE IN A MISMATCH  
 FT MUTAGEN 326 326 I->V: FULLY FUNCTIONAL IN A MISMATCH  
 FT MUTAGEN 326 326 O->L: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 552 552 ASSAY: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 672 672 R->P: DEFECTIVE IN A MISMATCH REPAIR  
 FT MUTAGEN 694 694 A->T: FULLY FUNCTIONAL IN A MISMATCH  
 FT MUTAGEN 258 258 P->L (IN REF. 1).  
 FT MUTAGEN 288 288 N->F (IN REF. 1).  
 FT MUTAGEN 708 708 S->L (IN REF. 1).  
 SQ SEQUENCE 769 AA; 87062 MW; B2DBB31DB3943171 CRC64;

Query Match 6.2%; Score 110; DB 1; Length 769;  
 Best Local Similarity 20.1%; Pred. No. 4.2;

Matches 63; Conservative 58; Mismatches 125; Indels 68; Gaps 11;

QY 13 NENGS---VDIAREMILKAKAGNNAVFCQFKADKLISALAPRAEVOIKNTGLESQ 68  
 DB 165 SHADYSKLDVAGYRAHSHKDIGSCCKF---GDSNYS-LSVPSSTVQ----- 210  
 QY 69 EMTKLEMKYDDYLHEKEVAAGSLNDFSTPDED---SIDFLASLKQKIMKIPSGEL 124  
 DB 211 -----DIRITVFNKSVASNLITFIHSKVEDNLNESVD-----GRVC 246  
 QY 125 NIPLYEAKADPIPKKIIISTGMATIDIKOSVSIFINNKPVGNITITHCETTPYPPF 184  
 DB 247 NINFTSKSKISPI-----PFINNRLVTCDLIRALMSVSNVLPKKNRPPIYIGIVIDPA 302  
 QY 185 EDVNIANAIIDLKHPKNNIGFSDH-----SSGFYALIAVPGIPIETIEGFTLDKSM 237  
 DB 303 VDENV-----HPTKREVRLSQDETIEKIANQHLALSALDSRTTKASISITNDE 354  
 QY 238 SGPDLIASIEPDELKALCTGVRCEKRS-LGSNSKVTFTASERKOKIYAKSIIAK--TRIK 294  
 DB 355 SLIPFDLTIESDBNRKSLRQAGVENSYYTANSQLEKAKROEKVLRIDASQAKITSPLS 414  
 QY 295 KGBVPSERKNTTKR 308  
 DB 415 SSGQFNFECSSTKR 428

RESULT 6  
 ID ISPG\_CHLMU STANDARD; PRT; 601 AA.  
 AC Q9PRT3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (EC 1.17.4.3) (1-  
 DE hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase).  
 GN ISPG OR TC0327.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mopn / N199;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Ultebrack T., Berry K., Bes S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumonise AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 CC -I- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,acpp) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (By similarity).  
 CC -I- CATALYTIC ACTIVITY: (E)-4-hydroxy-3-methylbut-2-en-1-yl  
 CC diphosphate + H(2)O + protein-dithiol = 2-C-methyl-D-erythritol  
 CC 2,4-cyclodiphosphate + protein-dithiol.  
 CC -I- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; sixth step.  
 CC -I- SIMILARITY: Belongs to the ispg family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE002300; AAF39191.1; -  
 CC PIR: C81715, C81715.  
 DR TIGR; TC0327;  
 DR HAMAP; MF\_00159; -; 1.  
 DR InterPro; IPR004588; ispg.  
 DR Pfam; PF04551; Gcpr; 1.  
 DR TIGRPFAMs; TIGR00612; ispg\_gcpr; 1.  
 KW Isoprene biosynthesis; Complete proteome; Oxidoreductase.  
 SQ SEQUENCE 601 AA; 67106 MW; CFA673DB9DD1E2B4 CRC64;

Query Match 6.1%; Score 109; DB 1; Length 601;  
 Best Local Similarity 20.0%; Pred. No. 3.6;

Matches 81; Conservative 57; Mismatches 122; Indels 146; Gaps 19;

QY 4 IYIABIGCNNGVDIAREMILKAKAGV-NAVFCQFKADKLISALA-----PKAEYQ 57  
 DB 50 IYALQECG-----EIVRTVQGLKEVAGCEQKDRIVQGNVITPLVADHFFPQAIIH 103  
 QY 58 IKN-----TSELDSQLEMTKLEKQDYHL----- 84  
 DB 104 VADFDKVRIRPGNVYDKRNNFSGKITYSDEQYTSLELTFKFPFLVAKCRKLRGWRIG 163  
 QY 85 -----KEXA-VSLNLPVSTPFDSDSID-----FL 108  
 DB 164 VNHGSLSERINQRYGDTIEGVAFSALETAAYCVNNDYKNIYFSKSSNPRFMVAAYRLA 223  
 QY 109 ASLKQIKWIPSGELANPLYEAKIADPIPKKIIISTGMAT-----DEIKOSVSIFIT 162  
 DB 224 RELDQRKVIY-----LHLGVTEAGSGM---DGMKISVSGICTLISEGIGDTIRCSITGSP 276

QY 163 NKKVAVGNITILECNTYEP-----TPEDVNLNAI--NDLKHPEKXNIGSDHSGEY 214  
 DB 277 TLEIIP-CDLKLKTAKESTKKNVPEIYHSKOLITOTTRKHPVEV-----YGLT 329  
 QY 215 AIAIAVPYGTIFERKAPFLDKSMGPDHLASIEPDELKLCIGVR-----C 260  
 DB 330 V-----SHT-----NDHLITTEPNTLO-CIGVDITTRKDLTPDGVV 367  
 QY 261 VEKLSGSKVATASERKXKIVAKSIIAKTEIKKGEVSEKXNITT 306  
 DB 368 IPKSMMS-SATVSEIKHARIFPKEDAPILNEMNELMNSDITLAI 412  
 RESULT 7  
 YG38 YEAST STANDARD; PRT; 713 AA.  
 AC P53276;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 80.2 kDa protein in ASN2-PHB1 intergenic region.  
 GN YGR128C.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talla E.,  
 RA Nawrocki A., del Bino S., Goffeau A.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; Z72913; CAA97140.1; -  
 DR PIR; S64437; S64437.  
 DR Germline; 141440; -  
 DR SGD; S0003360; YGR128C.  
 DR GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.  
 DR GO; GO:0030515; R:snRNA binding; IPI.  
 DR GO; GO:0030490; P:processing of 20S pre-rRNA; IMP.  
 DR KM Hypothetical protein; Transmembrane.  
 FT TRANSMEM 635 655 POTENTIAL.  
 SQ SEQUENCE 713 AA; 80190 MW; 6F5AE1492F841E49 CRC64;  
 Query Match 6.0%; Score 107.5; DB 1; Length 713;  
 Best Local Similarity 19.5%; Pred. No. 5.6; Tracts 115; Gaps 18;  
 Matches 76; Conservative 69; Mismatches 130; Indels 115; Gaps 18;  
 QY 14 HNGSVDIAREMILAKAGAVNAVFOFKADKLISAIAPAEVQIKNGELSGLEMTTK 73  
 DB 157 YTSDEFLSQQYKIKAK-AKWSIKIDT--KNSLIVAILQNGLIEI-----FPKLTILMS 208  
 QY 74 LEKQDYDLHMEYANSINDVF-----STPEPDSIDFLA 109  
 DB 209 FDISYNLAKAKFTENGTEYVVLCPLODDKCYKLELTDCGSGSSPIKLSSTIIE 268  
 QY 110 SLK-----QIKWIPSGE--LTLNPF--LEKIAKLPIDKKIISITGMATIDEIK 155  
 DB 269 GPFENSKCYGPGKLYKLNQGIYIYSPHCOLQGYTEPRWDK--LSPGDILISFOP 325  
 QY 156 QASV---IFINNAVPVGNITILKNT-----EYPPFEVNLNAINDLKHPKXNIGF 206  
 DB 326 VAVNRLVLTANNVYL--LDLHCSLTLSQRELTHTVTKFOLKSAVINSEKSNK----- 378  
 QY 207 SDHSSGFYALIAVPYGTIFERKAPFLDKSMGPDHLASIEPDELKLCIGVRCEVSKIG 266

DB 379 -----TAIGIS-----TNGENFTSSL--EIIINDVGTNTLKSLG 413  
 QY 267 SNSKVATASERKXKIVAKSIIAKTEIKKGEVSEKXNITT-----PENGISPM----- 316  
 DB 414 KSPGVGNND-----SSVILK-----FLPDKDINDRKVKNVDVSGSSVPLHGN 458  
 QY 317 EWNILKIAKADPFLPDELIIHSEPKNDE 346  
 DB 459 EVIERKSLAQNDITSPDIFPKELKIKEE 488  
 RESULT 8  
 NDM1 YEAST STANDARD; PRT; 2748 AA.  
 ID NDM1\_YEAST  
 AC Q00402; Q03767;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Nuclear migration protein NDM1.  
 GN NDM1 OR YDR150W OR YD9358.06.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=ATCC 28363 / FL100;  
 RA MEDLINE=92079507; PubMed=1145235;  
 RA Kornanev J., Kuentzel H.;  
 RA Perecko D., Kuentzel H.;  
 RT "Nuclear migration in 313 kDa NDM1 protein";  
 RL Mol. Gen. Genet. 230:277-287(1991).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=ATCC 28363 / FL100;  
 CC MEDLINE=92079507; PubMed=9169667;  
 CC Jacc C., Alt-Moerbe J., Andre B., Arnold W., Bahr A., Ballestra J.P.G.,  
 CC Barques M., Baron L., Becker A., Bileau N., Bloecker H., Blugnon C.,  
 CC Boskovic J., Brandt P., Bruckner M., Buitrago M.J., Coster F.,  
 CC Delaveau T., del Rey F., Dujon B., Fide L.G., Garcia-Carralajo J.M.,  
 CC Goffeau A., Gomez-Pertis A., Granclier C., Henemann V., Hankeln T.,  
 CC Hottelisel J.D., Jaeger W., Jimenez A., Jomiaux J.-L., Kremer C.,  
 CC Kuester H., Laamann P., Legros Y., Louis E.J., Koeller-Eleker S.,  
 CC Monnet A., Moro M., Mueller-Auer S., Nussbaumer B., Partido N.,  
 CC Paulin L., Perez J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,  
 CC Rieger M., Purnelle B., Rasmussen S.W., Remacha M., Revuelta J.L.,  
 CC Salmeron M., Salmeron D., Salmeron H.P., Saren A.M., Schaefer M.,  
 CC Schaefer M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,  
 CC Uribeastazu L.A., Verhaesele P., Vissers S., Voet M., Voickert G.,  
 CC Wagner G., Wandt R., Wedler S., Wedler S., Weill S., Harris D.E.,  
 CC Bowman S., Brown D., Churcher C.M., Connor R., Dedman K., Gentiles S.,  
 CC Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,  
 CC Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,  
 CC Walsh S.V., Barrall B.G., Dietrich F.S., Mulligan J.T., Allen E.,  
 CC Araujo R., Aviles E., Berro A., Carpenter J., Chen E., Cherry J.M.,  
 CC Chung E., Duncan M., Hunkle-Smith S., Hyman R.W., Kemp C.,  
 CC Lasakari D., Lew H., Lin D., Mosedale D., Nakatani K., Namath A.,  
 CC Oeffner P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,  
 CC Shogren T., Shroff N., Winant A., Yellon M.A., Botstein D.,  
 CC Davis R.W., Johnston M., Andrews S., Brinkman R., Cooper J., Ding H.,  
 CC Du Z., Pavllo A., Pullon L., Gattung S., Greco T., Hallsworth K.,  
 CC Hawkins J., Hillier L.W., Jier M., Johnson D., Johnson L.,  
 CC Kirzen S., Kucaba T., Langston Y., Latreille P., Le T., Mardis E.,  
 CC Meneses S., Miller N., Nham M., Pauley A., Peluso D., Rikkin L.,  
 CC Riles L., Tatch A., Trevarick E., Vignati D., Wilcox L., Wohldmann P.,  
 CC Vaudin M., Wilson R., Waterston R., Albertman K., Hand J., Heumann K.,  
 CC Kleene K., Mewes H.-W., Zollner A., Zaccaria P.;  
 CC "The nucleotide sequence of Saccharomyces cerevisiae chromosome IV";  
 CC Nature 387:75-78(1997).  
 CC -I- FUNCTION: Controls nuclear migration. NDM1 specifically controls  
 CC the interaction of the bud neck cytoskeleton with the pre-  
 CC divisional G2 nucleus perhaps by recognizing G2-specific



cytoplasmic microtubuli or other components of the nuclear envelope.

MISCELLANEOUS: Additional regions of lower homology to the repeat consensus (always starting with proline) are found in both flanking domains of the tandem repeats.

1- SIMILARITY: Contains 1 PH domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; X61236; CAA43554.1; .  
 EMBL; Z50046; CAA90372.1; .  
 GenBank; U140641; .

SGD; S0002557; NMI.

GO; GO:0005934; C:bud tip; IDA.  
 GO; GO:0005938; C:cell cortex; IDA.  
 GO; GO:0015631; F:tubulin binding; IPI.  
 GO; GO:0000226; F:microtubule cytoskeleton organization and b. . . . IMP.  
 GO; GO:0000065; P:nuclear migration (sensu Saccharomyces); IMP.  
 InterPro; IPR005529; DUF321.  
 InterPro; IPR001849; PH.  
 Pfam; PF03778; DUF321; 13.  
 Pfam; PF00169; PH; 1.  
 SMART; SM00233; PH; 1.  
 PROSITE; PS50003; PH\_DOMAIN; 1.

Repeat.

DOMAIN	593	1384	12.5 X TANDem REPEATS.
REPEAT	593	656	1.
REPEAT	657	727	2.
REPEAT	728	798	3.
REPEAT	799	862	4.
REPEAT	863	926	5.
REPEAT	927	990	6.
REPEAT	991	1054	7.
REPEAT	1055	1118	8.
REPEAT	1119	1182	9.
REPEAT	1183	1246	10.
REPEAT	1247	1310	11.
REPEAT	1311	1374	12.
REPEAT	1375	1384	13 (INCOMPLETE).
DOMAIN	2573	2683	PH.
CONFLICT	1570	1570	A->V (IN REF. 1).
CONFLICT	1822	1822	E->K (IN REF. 1).
CONFLICT	1960	1962	RAS->RHU (IN REF. 1).
CONFLICT	1971	1972	RD->RN (IN REF. 1).
CONFLICT	2049	2049	S->N (IN REF. 1).
CONFLICT	2637	2637	V->A (IN REF. 1).
SEQUENCE	2748 AA;	313030 MW;	EB4848E950621142 CRC64;

Query Match 6.0%; Score 107.5; DB 1; Length 2748;  
 Best Local Similarity 21.1%; Pred. No. 30;  
 Matches 80; Conservative 61; Mismatches 113; Indels 125; Gaps 19;

18 VDIARE-----MILAKKAGNAVAVKQFTKAKLISALIPKAEYQIKNGELSGLEMTKK 73  
 1871 VAISKDYELVNAKREKPOLYLLKISSKIDHT--VPLSEYVNAVYNNPNSLSLUKE 1927

74 -----LEMKYDY-----LHNAEYAVSLN-----LDVEST--PREDESID 106

1928 KAVLNHLLIKEDDYKNILAVSEHPVHLSEKXSLINKVLVDODDFATMSRSLKPTID 1987

107 FLASIKKQIKWIPGSE-----LHNPYLEKAKLPIPKKIIS-----TGM 148

1988 FLSTALSMGKILVNESTHKNKLSPEDEFLTKAKKQGLIISSEKYSERLDQIDR 2047

149 AIIIEIKOSVSIPIFNKKVPGNITILH--CNTEYPTP--PEDVNLNAINDLKKHPPKNI 204  
 2048 PSLDTLKKKKAIF--DSIIVENIEEQGLVNTTSCPPITTEDLKVYA----- 2092

205 GFDSSGGFYAAIAVYGITFIKFTPLDKSMGPDHLASIEPDELKHLICIVRCYKS 264  
 2093 -----HKGMELCLOKPKKKSABAE-----RIDEQS 2120

265 LG--SNKVVTA-----ERKKIVAKS-----IIKTE-IKKGEV--SKNI 304  
 2121 INTTSSSTTSSKFTALDNDNIELNRELONNEDYDIIKSKSTVXDITIFPAVENI 2180

305 --TTKRPNGISPEWNTL 321  
 2181 KNSAEKIGYLVPEKSN 2199

RESULT 9  
 YOPM\_YERPE STANDARD; PRT; 409 AA.  
 ID YOPM\_YERPE  
 AC P1778; O68701;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Outer membrane protein yopM.  
 GN YOPM OR YOP48 OR YPCD1.26C OR Y5054 OR Y0059.  
 OS Yersinia pestis.  
 OG Plasmid pCD1.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5 / Biovar Mediaevalis;  
 RX MEDLINE=89359090; PubMed=2670888;  
 RA Leung K.Y., Straley S.C.;  
 RT "The yopM gene of Yersinia pestis encodes a released protein having  
 homology with the human platelet surface protein GpIIb alpha.";  
 J. Bacteriol. 171:4623-4632(1989).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5 / Biovar Mediaevalis;  
 RX MEDLINE=98427122; PubMed=9746557;  
 RA Perry R.D., Straley S.C., Fetherston J.D., Rose D.J., Gregor J.,  
 RA Blatter F.R.;  
 RT "DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of  
 Yersinia pestis KIM5.";  
 J. Infect. Immun. 66:4611-4623(1998).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5 / Biovar Mediaevalis;  
 RX MEDLINE=98422474; PubMed=9748454;  
 RA Hu P., Elliott J., McCreedy P., Skowronski E., Garmes J.,  
 RA Kobayashi A., Brubaker R.R., Garcia B.;  
 RT "Structural organization of virulence-associated plasmids of Yersinia  
 pestis.";  
 J. Bacteriol. 180:5192-5202(1998).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Partholl J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baser S., Besman D., Bentley S.D., Brooks K., Cedeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies K.M., Davis F., Dougan G.,  
 RA Fellwell T., Hamlin N., Holroyd S., Jagsels K., Kariyasev A.V.,  
 RA Leather S., Moulis S., Oyston P.C.F., Quail M.A., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 Nature 413:523-527(2001).  
 CC -1- STRUCTURAL LOCATION: OUTER MEMBRANE AND SECRETED.  
 CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.  
 CC -1- SIMILARITY: TO S.FLEXNERI IPAH 7.8.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF074612; AAC69806.1; -  
 DR EMBL; AF053946; AAC62580.1; -  
 DR EMBL; AL117189; CAB54903.1; -  
 DR PIR; A33950; A33950.  
 DR PIR; T43599; T43599.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR007088; LRR Dact.  
 DR Pfam; PF00560; LRR; 7.  
 KM Outer membrane; Repeat; leucine-rich repeat; Antigen; Plasmid;  
 KM Complete proteome.

FT REPEAT 69 90 LRR 1.  
 FT REPEAT 91 110 LRR 2.  
 FT REPEAT 111 129 LRR 3.  
 FT REPEAT 130 152 LRR 4.  
 FT REPEAT 153 177 LRR 5.  
 FT REPEAT 179 194 LRR 6.  
 FT REPEAT 195 217 LRR 7.  
 FT REPEAT 236 256 LRR 8.  
 FT REPEAT 257 281 LRR 9.  
 FT REPEAT 296 316 LRR 10.  
 FT REPEAT 317 341 LRR 11.  
 FT REPEAT 356 381 LRR 12.  
 FT CONFLICT 142 183 MISSING (IN REF. 1).  
 SQ SEQUENCE 409 AA; 46203 MW; E22400C671A59951 CRC64;

Query Match 6.0%; Score 107; DB 1; Length 409;  
 Best Local Similarity 20.3%; Pred. No. 3.1; Matches 53; Conservative 63; Mismatches 97; Indels 48; Gaps 12;

CC 45 KLISALAPKAEYQIKNTGSELSELEMTKLEKDYDYLHMEYAVSLNDFSTPEDS 104  
 DB 124 KALSDLEPLEY-----LGVNNQLE--KLPELQSSFLKIIDV--NNSLKLPPLPS 174  
 QY 105 IDPLASLKQKIMIPSEELINLY-----LEKIANLPIDPKKIISTGMATIDEI 154  
 DB 175 LELIAGNNQLELP--ELQNLPELTAIYADNNSLKLPPLPSISIV--AGNNILSEH 230  
 QY 155 KQVSIFINKKVPVGNITLHCTEYPTPEVDNL--NAINDLKKEPKKNIIGFSDHSG 212  
 DB 231 PELQNLPELTIYADN--NLKTLPELPPSLHAIYADNNTLDFE--LPO--SLTFIDVSEN 287  
 QY 213 FYAIAAVPYGIFTEKHFPLDKWSPDHLSIEPDELKHLCTIGVCVEKSIQSNRYV 272  
 DB 288 IFSGLSLEPNNLYL-----NASSNEIRSLCDLPPLSEILAVSNKKLI 330  
 QY 273 ---TASERKKIVARKSIIX 290  
 DB 331 ELPALEPRLERLIASRNHIAE 351

RESULT 10  
 YA84 SCHPO STANDARD; PRT; 1428 AA.  
 ID YA84 SCHPO  
 AC 009773;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C22F3.04 in chromosome I.  
 GN SPAC22F3.04.  
 OS Schizosaccharomyces pombe (Fission Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Galliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble S., Hamilton N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkhardt G., Aert R., Robben D., Grymoprez B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Pritzc C., Holzer E., Moesti D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leinrich R., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Mamburt R., Purrelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe".  
 RL Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -1- SIMILARITY: TO S.POMBE SPAC56F8.02.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; 254285; CA91069.1; -  
 DR PIR; T38187; S62419.  
 DR GeneDB; SPAC22F3.04; -  
 DR InterPro; IPR000873; AMP-bind.  
 DR Pfam; PF00501; AMP-binding; 3.  
 KM Hypothetical protein; Transmembrane.  
 FT TRANSMEM 1028 1048 POTENTIAL.  
 FT TRANSMEM 1303 1323 POTENTIAL.  
 FT TRANSMEM 1333 1353 POTENTIAL.  
 SQ SEQUENCE 1428 AA; 162394 MW; A8BD03AE0A9E9899 CRC64;

Query Match 5.9%; Score 105.5; DB 1; Length 1428;  
 Best Local Similarity 23.5%; Pred. No. 18; Matches 79; Conservative 44; Mismatches 144; Indels 69; Gaps 14;

CC 14 HNGSYDIRPNT-----LKAERGV-NAYVGTETKADKLIALAPKAEYQIKNTGE 63  
 DB 785 HONKVSKEPENTLNEFTLIDFLKTSARTPTSPALITFNALEKVEILTWHAFELKVASH 844  
 QY 64 LE-SOLENTYK-----LEMKYDDYLHMEYAVSLNDFSTPEDSIDPLASLKQKIM 117  
 DB 845 VEYQGLSVAKARSHILLVYDPLEFLSIHSCFHLGVIPIFQIREISQILGEIEEFLK 904  
 QY 118 IPSG-----ELNLP-----YLEKIANLPIDPKKIISTGMATIDEIKQVS 159  
 DB 905 IAKAPVAVAILVDHVKVLSKSRDISNHFQCTCIDLNKAPKPIETFTISPISKRAKALN 964  
 QY 160 IF-----INNKVPVGNITL--LHCTEYPTPEVDNLAINDKKIFPNN-----IGSD 208  
 DB 965 LITPELANKKGVMLISINKLEDSIIPTQSHSLAFCHQKEFVIGNEEKPIIGIE 1024  
 QY 209 HSSG---FYAIAAVPYGI--TFIEKHFPLDKWSPDHLSIEPDELKHLCTIGVCVEKS 264  
 DB 1025 FSSGQGLHTALTALGVYAVPTLLIKQENTL--CNNGSILFPAIEQNSLSKVMIPANI 1082

QY 265 L-----GNSKRVTA-----SERKNTV 282  
DB 1083 FSTAGGNSIYINSTESSIIVPCYDRPISSRVNII 1118

RESULT 11  
ID YFE9\_YEAST STANDARD; PRT; 623 AA.  
AC P43554;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Hypothetical 70.3 kDa protein in ALK2-EMP4 intergenic region.  
GN YF1049W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
[1]  
[1] SEQUENCE FROM N.A.  
RA STRAIN=288C / AB972;  
RA MEDLINE=95400292; PubMed=7670463;  
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
RA Yamazaki M., Tashiro H., Eki T.;  
RT Analysis of the nucleotide sequence of chromosome VI from  
RT Saccharomyces cerevisiae.  
RL Nat. Genet. 10:261-268(1995).  
CC -1- SIMILARITY: TO YEAST NPL6.  
CC -1-  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D50617; BAA09192.1; -  
DR PIR: S56206; S56206;  
DR GeneOnline: 140107;  
DR SGD: S0001845; YFL049W.  
DR Hypothetical protein.  
SQ SEQUENCE 623 AA; 70275 MW; A8AC00CA8F0ED0E CRC64;

Query Match 5.8%; Score 103.5; DB 1; Length 623;  
Best Local Similarity 22.3%; Pred. No. 8.9;  
Matches 86; Conservative 41; Mismatches 134; Indels 125; Gaps 19;

QY 9 EIGNGNSVDIARE-----MILKAKGVNAVKQTFKADKLISALPKAEYQIKNT 61  
DB 254 EISANTNTCTISKFSQSPYIETQPSAIRAATYENFKGSHISAIVP--GQSISGT 311  
QY 62 CELESQLEMKKLEMKYDDVILHKEVAVSLNDVFSPPEDSDIDFLASKQKIMKIPSG 121  
DB 312 LEISAQ-----FVPRYHKNKSFQALQAMANDPIG 343  
QY 122 ELANLPILEKIAKL--PIPKKIIISTGMAIDIKOSVSIFFNNKPVQNT--TILHGN 177  
DB 344 R-----HEBLLAQYESQAPD-----GSASISLPNHIPSVNPK--PIKMLSSILDIN 390  
QY 178 TEYTPPEDVNAIINDLKHPKPNQNGF-----SHSGG--FYAIAAIP 221  
DB 391 VSSSKNKKSEBENEMIKWNGQKHNNQSLNNGKFSLLPKSKXENSGKQOYR----- 444  
QY 222 YGIFIKKFTLDSMGPDHIASIEPDELK--HLCIGVCYKESIGSNKSVT----- 273  
DB 445 -GLPLVEKNITLER-----LKQLTNEIKELHLDAV--FVYTGLOVRAKRTKKKK 495  
QY 274 -----ASERNKIVAKRSIATKTEIKGVFSE-----KNITTKRGNG 312  
DB 496 YWQYKAGIPIGLKRSQIDEPKRYL--KDVLAQTSVTN--FNEITWDTBITTKRVNPN 551

QY 313 ISPMENYELGKIAEDPIDELIIH 338  
DB 552 -----NELGNCNIDFKPPIYSH 570

RESULT 12  
ID PYCB\_METUA STANDARD; PRT; 567 AA.  
AC O58628;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Pyruvate carboxylase subunit B (EC 6.4.1.1) (Pyruvic carboxylase B).  
GN PYCB OR MJ1231.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
[1]  
[1] SEQUENCE FROM N.A.  
RA STRAIN=JUL-1 / DSM 2661 / ATCC 43067;  
RA MEDLINE=96337999; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Nguyen D.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
RA Uetrecht J.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurt M.A., Kane B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RT Science 273:1058-1073(1996).  
RL [2]  
RP SEQUENCE OF 190-125, 260-270, 277-289, 277-289, 309-325, 328-358;  
RP 370-380, 386-409, 422-438, 491-506 AND 491-506, AND FUNCTION.  
RX MEDLINE=2104791; PubMed=1195096;  
RA Mukhopadhyay B., Patel V.J., Wolfe R.S.;  
RT "A stable archaeal pyruvate carboxylase from the hyperthermophile  
RT Methanococcus jannaschii.";  
RT Arch. Microbiol. 174:406-414(2000).  
RL  
CC -1- FUNCTION: Pyruvate carboxylase catalyzes a 2-step reaction,  
CC involving the ATP-dependent carboxylation of the covalently  
CC attached biotin in the first step and the transfer of the carboxyl  
CC group to pyruvate in the second.  
CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3) (-) = ADP + phosphate +  
CC oxaloacetate.  
CC -1- COPACTOR: ATP, magnesium (or manganese or cobalt), pyruvate and  
CC bicarbonate.  
CC -1- ENZYME REGULATION: INHIBITED BY MAGNESIUM, WHEN ITS CONCENTRATION  
CC EXCEEDED THE ATP ONE, AND BY HIGH CONCENTRATION OF ATP AND ALPHA-  
CC KETOGLUTARATE.  
CC -1- PATHWAY: Gluconeogenesis.  
CC -1- SUBUNIT: Heterooctamer of four A and four B subunits.  
CC -1- MASS SPECTROMETRY: MW=64160, METHOD=VALDIT.  
CC -1- WISCELLANEOUS: ITS OPTIMUM PH IS 8.5 AND THE OPTIMUM TEMPERATURE  
CC IS 80-90 DEGREES CELSIUS.  
CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES  
CC AND CARBAMYL PHOSPHATE SYNTHETASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U67563; AAB99233.1; -  
DR PIR: P64453; P64453.  
DR HSP: P02905; 1A6X.  
DR TIGR: MJ1231;  
DR InterPro: IPR001882; Biotin\_BS.

DR InterPro: IPR000089; Biotin\_lipoyl.  
 DR InterPro: IPR000891; HMGCL-like.  
 DR InterPro: IPR005776; OadA.  
 DR InterPro: IPR003379; PYC OAD.  
 DR Pfam: PF00364; biotin\_lipoyl; 1.  
 DR Pfam: PF00682; HMGCL-like; 1.  
 DR Pfam: PF02436; PYC OAD; 1.  
 DR TIGRFAMs: TIGR01108; OadA; 1.  
 DR PROSITE: PS00188; BIOTIN; 1.  
 KW Ligase; Multifunctional enzyme; Glucanogenesis; Magnesium; Pyruvate;  
 KW Biotin; Complete proteome.  
 FT BINDING 53 53 BIOTIN (BY SIMILARITY).  
 FT BINDING 53 53  
 SQ SEQUENCE 567 AA; 63907 KM; 5E07800622545628 CRC64;

Query Match 5.8%; Score 102.5; DB 1; Length 567;  
 Best Local Similarity 19.7%; Pred. No. 9.3;  
 Matches 75; Conservative 71; Mismatches 137; Indels 97; Gaps 19;

QY 10 ICGNNGSDIAREMLKAKKGVAKVQT-----KADKLISALAPKAE 55  
 DB 86 VGRHYPD-DIVEKFKAKHENGID--FRIDALNDVNNMTAKTKVGAEEVQAIC 142  
 QY 56 YQIKNTGESOLEMTKLEMYKDYHMEYAVSINDVSPEDSDIDELASIKQKI 115  
 DB 143 YIISPHTIDQVELAKLEEMGCCSICKMAAGL-----TPY--EGEYLVKRLKEI 194  
 QY 116 WKIP-----SGELINDYLEKIAKLPIDKKIITSTGMATIDIKOSVSIPT--NNKV 166  
 DB 195 -SLPIDVHSHCTSGALPMTYLVKIE-----AADWVDCASIFPANGTSP 238  
 QY 167 PVGNITILHONTPEYPPFEDVNLAINDLKKAP---PKNIGESHSNGFAALAA--V 220  
 DB 239 PFSIIVALKGTITDT--GLDILNEDYDMKREKRYLFSISQIVARVIVYV 295  
 QY 221 PVG-----ITFEKHTLDRKMSGGPHLASIEPDELKELCTG---VRCVEKSLGNSKV 271  
 DB 296 PGMLSNLVSQLKEQALDKFESEVQDEIPIVRVD-----LGYPLVPTISQIVGQAVL 349  
 QY 272 VFSSEKNTIVAKSI-----IAKTEIKGVSEKNTITKPPANGIS 314  
 DB 350 NVLTERKYKITNEVNVYKGPVKKPPAPINPELKVULDEG---EKETTC-REAP-L 403  
 QY 315 PWEYMLGKIAEDPIPDE 334  
 DB 404 PPEWEVYKKEAEKGIYKKE 423

RESULT 13  
 VCAL\_MOUSE STANDARD; PRT; 739 AA.  
 AC P29533;  
 DE 01-APR-1993 (Rel. 25, Created)  
 DE 01-APR-1993 (Rel. 25, Last sequence update)  
 DE 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1).  
 GN VCAM1 OR VCAM-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=FVB; TISSUE=Lung;  
 RA MEDLINE=92181437; PubMed=1371918;  
 RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wyck M.,  
 RA Burkly L., Miyake K., Kincaid P., Lobb R.;  
 RT Cloning of murine and rat vascular cell adhesion molecule-1.;  
 RL Biochem. Biophys. Res. Commun. 183:163-169(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Lymph node;  
 RA MEDLINE=93246254; PubMed=7683304;

RA Ataki M., Ataki K., Vassalli P.; VCAM-1 cDNA."  
 RT "Cloning and sequencing of mouse VCAM-1 cDNA."  
 RL Gene 126:261-264(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=129; TISSUE=Embryo;  
 RX MEDLINE=94117008; PubMed=7507076;  
 RA Cypulsky M.I., Allen-McCann M., Collins T.;  
 RT "Structure of the murine VCAM1 gene."  
 RL Genomics 18:387-391(1993).  
 RN [4]  
 RP SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).  
 RC STRAIN=NIH Swiss; and 129/Sv;  
 RA Kuner A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M. Jr.;  
 RA Ballantyne C.M.; VCAM-1 to the EMBL/GenBank/DBJ databases.  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=FVB; TISSUE=Lung;  
 RX MEDLINE=93232042; PubMed=7682556;  
 RA Moy P., Lobb R., Tizard R., Olson D., Hession C.;  
 RT "Cloning of an inflammation-specific phosphatidyl inositol-linked  
 form of murine vascular cell adhesion molecule-1."  
 RL J. Biol. Chem. 268:8835-8841(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RX MEDLINE=95015899; PubMed=7523515;  
 RA Kuner A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M.,  
 RA Ballantyne C.M.; Molecular cloning, mapping, and analysis of a  
 truncated form."  
 RL J. Immunol. 153:4088-4098(1994).  
 RN [7]  
 RP SEQUENCE OF 311-345 FROM N.A. (ISOFORM 2).  
 RC STRAIN=FVB/N; TISSUE=Kidney;  
 RX MEDLINE=93317595; PubMed=7687058;  
 RA Terry R.W., Kwee L., Levine J.P., Labov M.A.;  
 RT "Cytokine induction of an alternatively spliced murine vascular cell  
 adhesion molecule (VCAM) mRNA encoding a  
 glycosylphosphatidylinositol-anchored VCAM protein."  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5919-5923(1993).  
 RN [8]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RC TISSUE=endothelial cells;  
 RA Korenaga R., Ando J., Tsuboi H., Kamiya A.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION  
 IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1  
 INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL  
 TRANSDUCTION. THE VCAM1/VIA4 INTERACTION MAY PLAY A  
 PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE  
 EMIGRATION TO SITES OF INFLAMMATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG ISOFORM) OR  
 ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (SHORT ISOFORM).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=Long;  
 CC IsoId=P29533-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Short;  
 CC IsoId=P29533-2; Sequence=VSP\_002581, VSP\_002582;  
 CC TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as  
 well as on macrophage-like and dendritic cell types in both normal  
 and inflamed tissue.  
 CC -1- PTM: The GPI-anchor is located on position 319 of isoform 2.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: May 6, 2004, 08:59:19 ; Search time 33.8671 Seconds  
(without alignments)  
3223.466 Million cell updates/sec

Title: US-09-930-440b-8

Perfect score: 1778

Sequence: 1 MSNIIYVAIEGNCNHSVDI.....EQDFIPELIIHSEFKNGR 346

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioplasmid:\*  
17: sp\_archaeoplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1778	100.0	346	2	Q46675
2	1239	69.7	346	2	Q93N02
3	981.5	55.2	341	2	Q87186
4	981.5	55.2	341	16	Q8E501
5	979.5	55.1	339	2	Q9ALM6
6	979.5	55.1	339	2	Q93T12
7	979.5	55.1	341	2	Q9RRC0
8	979.5	55.1	341	16	Q9BDE3
9	951.5	53.5	341	2	Q9AQ16
10	868	48.8	344	16	Q8F5Q8
11	867	48.8	334	16	Q9PMX2
12	844	47.5	334	2	Q7X523
13	769	43.3	333	16	Q7U911
14	723.5	40.7	333	16	Q8BDZ7
15	708.5	39.8	357	16	Q87770
16	676	38.0	338	2	Q9RDX5

17	651	36.6	361	2	Q8KX2	Q8KX2 pseudomonas
18	646	36.3	229	2	Q9E9F4	Q9E9F4 campylobact
19	642.5	36.1	356	16	Q89HJ8	Q89HJ8 bradyrhizob
20	534	30.0	351	16	Q893U6	Q893U6 clostridium
21	520.5	29.3	753	16	Q8F5Q4	Q8F5Q4 leptospira
22	504	28.3	359	11	Q94JH0	Q94JH0 mus musculus
23	502	28.2	359	11	Q94J77	Q94J77 mus musculus
24	500.5	28.1	343	2	Q93J08	Q93J08 campylobact
25	497.5	28.0	349	16	Q57265	Q57265 neisseria m
26	496.5	27.9	343	2	Q7X528	Q7X528 campylobact
27	495.5	27.9	641	16	Q89HL9	Q89HL9 bradyrhizob
28	493	27.7	350	16	Q97H26	Q97H26 clostridium
29	493	27.7	749	16	Q82UC3	Q82UC3 nitrosomonas
30	492.5	27.7	343	16	Q9PMY2	Q9PMY2 campylobact
31	484.5	27.2	346	2	Q9F0M8	Q9F0M8 campylobact
32	484.5	27.2	352	2	Q9R3S2	Q9R3S2 aeromonas p
33	481	27.1	351	2	Q8KH52	Q8KH52 pseudomonas
34	480.5	27.0	346	2	Q938X5	Q938X5 campylobact
35	479.5	27.0	346	16	Q8A711	Q8A711 bacteroides
36	478.5	26.9	346	2	Q93FV4	Q93FV4 campylobact
37	478.5	26.9	346	2	Q9LEQ4	Q9LEQ4 campylobact
38	478.5	26.9	346	2	Q93D04	Q93D04 campylobact
39	476.5	26.8	346	2	Q93MP9	Q93MP9 campylobact
40	475.5	26.7	346	2	Q9LAK2	Q9LAK2 campylobact
41	471.5	26.5	346	2	Q93C24	Q93C24 campylobact
42	467	26.3	349	16	Q7VH05	Q7VH05 helicobacter
43	465.5	26.2	327	2	Q52907	Q52907 campylobact
44	464	26.1	754	16	Q7H964	Q7H964 synechococc
45	456.5	25.7	354	2	Q9R3M1	Q9R3M1 rhizobium m

## ALIGNMENTS

RESULT 1  
ID Q46675 PRELIMINARY; PRT; 346 AA.  
AC Q46675;  
DT 01-NOV-1996 (T-REMBLrel. 01, Created)  
DT 01-NOV-1996 (T-REMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (T-REMBLrel. 24, Last annotation update)  
DE NEUB protein.  
GN NEUB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KI;  
RX MEDLINS=95113767; PubMed=7814319;  
RA Silver R.P., Wright L.F., Annunziato P.W.;  
RT "Nucleotide sequence and genetic analysis of the neub and neub genes  
in region 2 of the polystyolic acid gene cluster of Escherichia coli  
KI.";  
RT KI.\*;  
RL J. Bacteriol. 177:312-319 (1995).  
DR EMBL: U05248; AAC43302.1; --  
DR FTR; 169836; 169836.  
DR GO; GO:0016051; Polysaccharide biosynthesis; IRA.  
DR InterPro; IPR006014; Antifreeze dom.  
DR InterPro; IPR006190; Antifreeze-like.  
DR InterPro; IPR004144; Neub.  
DR Pfam; PF03154; Antifreeze; 1.  
DR Pfam; PF03152; Neub; 1.  
DR PROSITE; PS00844; AFP\_LIKE; 1.  
DR PROSITE; PS00844; AFP\_LIKE; 1.  
SQ SEQUENCE 346 AA; 38693 MW; EF558B8F9B15A02B CRC64;  
Query Match 100.0%; Score 1778; DB 2; Length 346;  
Best Local Similarity 100.0%; Pred. No. 2.4e-117;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MSNIIYVAIEGNCNHSVDIARMIKAYGAGNAYVGTFRADKLSIAIAPAEYQIKR 60  
|||||

Thu May 13 11:53:14 2004

us-09-930-440b-8.rsp

Page 2

Db 1 MSNIYVAEIGCHNGSVDIAREMILKAKAGVNAVQOTFKADKLISALAPRAEYOIKN 60  
QY 61 TGELESQLEMTKLEKEMKYDDYHMEYAVSLNDVSTPDESDSIDFLASKOKIMKIPS 120  
Db 61 TGELESQLEMTKLEKEMKYDDYHMEYAVSLNDVSTPDESDSIDFLASKOKIMKIPS 120  
QY 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKYVPVGNITILHONTY 180  
Db 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKYVPVGNITILHONTY 180  
QY 181 PTFEPEDVNAINDLKKHPPKNNIGFSDHSGFYAAIAAPVGIITFEKFTLIDKMSGP 240  
Db 181 PTFEPEDVNAINDLKKHPPKNNIGFSDHSGFYAAIAAPVGIITFEKFTLIDKMSGP 240  
QY 241 DHIASIEPDELKHLICIGVRCVKSIGNSKVVTASERKNKIIVAKSIIAKTEIKKGEVS 300  
Db 241 DHIASIEPDELKHLICIGVRCVKSIGNSKVVTASERKNKIIVAKSIIAKTEIKKGEVS 300  
QY 301 EKNTTRPGNGISPMEMWNLGKIAODFIPELIIHSEFNQGE 346  
Db 301 EKNTTRPGNGISPMEMWNLGKIAODFIPELIIHSEFNQGE 346

RESULT 2

Q93NQ2 PRELIMINARY; PRT; 346 AA.

AC Q93NQ2: 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE NAB.  
GN Bacteriophage phi.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2129197; PubMed=11404020;  
RA Wang L., Briggs C.E., Rothmund D., Fratantoni P., Luchansky J.B.,  
RT "Sequence of the E. coli O104 antigen gene cluster and identification  
of O104 specific genes."  
RL Gene 270:231-236 (2001).  
DR EMBL; AF61371; AAK64368.1; -  
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro; IPR006014; Antifreeze dom.  
DR InterPro; IPR006190; Antifreeze-like.  
DR Pfam; PF01354; Antifreeze; 1.  
DR Pfam; PF01302; Neut; 1.  
DR PROSITE; PS50844; AFP\_LIKE; 1.  
SQ SEQUENCE 346 AA; 35052 MW; C167A8A5A0A2721D CRC64;

Query Match 69.7%; Score 1239; DB 2; Length 346;  
Best Local Similarity 68.9%; Pred. No. 2.3e-79;  
Matches 237; Conservative 48; Mismatches 59; Indels 0; Gaps 0;

QY 1 MSNIYVAEIGCHNGSVDIAREMILKAKAGVNAVQOTFKADKLISALAPRAEYOIKN 60  
Db 1 MSNIYVAEIGCHNGSVDIAREMILKAKAGVNAVQOTFKADKLISALAPRAEYOIKN 60  
QY 61 TGELESQLEMTKLEKEMKYDDYHMEYAVSLNDVSTPDESDSIDFLASKOKIMKIPS 120  
Db 61 TGELESQLEMTKLEKEMKYDDYHMEYAVSLNDVSTPDESDSIDFLASKOKIMKIPS 120  
QY 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKYVPVGNITILHONTY 180  
Db 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKYVPVGNITILHONTY 180  
QY 181 PTFEPEDVNAINDLKKHPPKNNIGFSDHSGFYAAIAAPVGIITFEKFTLIDKMSGP 240  
Db 181 PTFEPEDVNAINDLKKHPPKNNIGFSDHSGFYAAIAAPVGIITFEKFTLIDKMSGP 240

QY 241 DHIASIEPDELKHLICIGVRCVKSIGNSKVVTASERKNKIIVAKSIIAKTEIKKGEVS 300  
Db 241 DHIASIEPDELKHLICIGVRCVKSIGNSKVVTASERKNKIIVAKSIIAKTEIKKGEVS 300  
QY 301 EKNTTRPGNGISPMEMWNLGKIAODFIPELIIHSEFNQGE 344  
Db 301 EKNTTRPGNGISPMEMWNLGKIAODFIPELIIHSEFNQGE 344

RESULT 3

Q87186 PRELIMINARY; PRT; 341 AA.

AC Q87186: 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE NEUB protein.  
GN NEUB.  
OS Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1311;  
RN  
RP SEQUENCE FROM N.A.  
RA Yamamoto S., Miyake K., Iijima S.;  
RT "Identification and Characterization of cps (capsular polysaccharide)  
genes from Streptococcus agalactiae Type Ia."  
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
RN  
RP SEQUENCE FROM N.A.  
RA Miyake K., Yamamoto S., Koike Y., Watanabe M., Iijima S.;  
RT "Molecular Characterization of Type-Specific Capsular Polysaccharide  
Biosynthesis Genes of Streptococcus agalactiae Type Ia."  
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB017355; BAA3753.1; -  
DR EMBL; AB028896; BAA82287.1; -  
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro; IPR006190; Antifreeze dom.  
DR InterPro; IPR006190; Antifreeze-like.  
DR Pfam; PF01354; Antifreeze; 1.  
DR Pfam; PF01302; Neut; 1.  
DR PROSITE; PS50844; AFP\_LIKE; 1.  
SQ SEQUENCE 341 AA; 38091 MW; 17620BF82A2BA340 CRC64;

Query Match 55.2%; Score 981.5; DB 2; Length 341;  
Best Local Similarity 56.3%; Pred. No. 3.2e-61;  
Matches 192; Conservative 55; Mismatches 89; Indels 5; Gaps 2;

QY 4 IYIYVAEIGCHNGSVDIAREMILKAKAGVNAVQOTFKADKLISALAPRAEYOIKN 63  
Db 4 IYIYVAEIGCHNGSVDIAREMILKAKAGVNAVQOTFKADKLISALAPRAEYOIKN 63  
QY 64 LBSQLEMTKLEKEMKYDDYHMEYAVSLNDVSTPDESDSIDFLASKOKIMKIPS 123  
Db 64 LBSQLEMTKLEKEMKYDDYHMEYAVSLNDVSTPDESDSIDFLASKOKIMKIPS 123  
QY 122 TULPFLBKIGK---QOKVILSTGVAWEEIHOAVNIRMGCTT--DISILHCTTEYPTP 176  
Db 122 TULPFLBKIGK---QOKVILSTGVAWEEIHOAVNIRMGCTT--DISILHCTTEYPTP 176  
QY 184 FEDVNAINDLKKHPPKNNIGFSDHSGFYAAIAAPVGIITFEKFTLIDKMSGP 243  
Db 184 FEDVNAINDLKKHPPKNNIGFSDHSGFYAAIAAPVGIITFEKFTLIDKMSGP 243  
QY 237 ASATPDIILALVKGRIYEGALGREKIPDPVEEKNKIIVAKSIIVAKSIIVAKSIIVAKSI 296  
Db 237 ASATPDIILALVKGRIYEGALGREKIPDPVEEKNKIIVAKSIIVAKSIIVAKSIIVAKSI 296  
QY 304 ITTRKPGNGISPMEMWNLGKIAODFIPELIIHSEFNQGE 344  
Db 304 ITTRKPGNGISPMEMWNLGKIAODFIPELIIHSEFNQGE 344



## RESULT 4

Q8E501 PRELIMINARY; PRT; 341 AA.  
 ID Q8E501  
 AC Q8E501;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 NEUB OR GBS1236.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 NCBI\_TaxID=216495;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Ruenick C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Masdek T., Zouine M., Couve E., Lalloué L., Poyart C., Tieu-Cuot P.,  
 RA Kunst F.;  
 RT Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL; AL766849; CAD46895.1; -  
 DR SAGelst; gbs1236; -  
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro; IPR006014; Antifreeze dom.  
 DR InterPro; IPR006190; Antifreeze \_like.  
 DR InterPro; IPR004144; Neug.  
 DR Pfam; PF01354; Antifreeze; 1.  
 DR Pfam; PF03102; Neug; 1.  
 DR PROSITE; PS50844; AFP\_LIKE; 1.  
 DR Hydrothermal protein; Complete proteome.  
 KW SEQUENCE 341 AA; 38091 MM; 176208F82BA340 CRC64;  
 SQ

Query Match 55.1%; Score 981.5; DB 16; Length 341;  
 Best Local Similarity 56.3%; Pred. No. 3.2e-61;  
 Matches 192; Conservative 55; Mismatches 89; Indels 5; Gaps 2;

QY 4 IYVAEIGCNHNSVDIAREMILKAEAGVNAVKQTFKADKLISALAPRAEYQKNTGE 63  
 DB 2 VYIIAEIGCNHNSVDIAREMILKAEAGVNAVKQTFKADKLISALAPRAEYQKNTGT 61  
 QY 64 LESQLEMTKLEMKYDYVILHMEYAVSLNDVSTPDESIDPLASIKOKIWKIPSGEL 123  
 DB 62 ADSQLEMTKLEMKYDYVILHMEYAVSLNDVSTPDESIDPLASIKOKIWKIPSGEI 121  
 QY 124 LNEPYLEKIAKLPIDPKKIIISTGMATIDEIKOSVSIINNKPVPNGNTIIHCNTEYPTP 183  
 DB 122 TNLPLEKIKG--QOKKVILSTGMAVMEIHOAVNIIIRONGTT--DISILHCTTEYPTP 176  
 QY 184 FEDVNLAINDIKKEFPKNNIGFSDHSSGFYAALAAVEYGTTFEKFHTLDSKSGPDHL 243  
 DB 177 YPSLNLAVIHTLKDFKDLITIGSDHSIGSEVPAAAGAEVIEKHTLDTNNMGDPDK 236  
 QY 244 ASIEDDELKHLCTIGRCYKESKSGNSKYVTASERKNTIVAKSIATKIEIKGAVPSKKN 303  
 DB 237 ASATPDIIAALVKGRIVEQALGRREKIPDPVEEKNTIVAKSVVALPKIKGDIYSTEN 296  
 QY 304 ITTKRPNNGISPMEWYNLGLKIAEODFIPDELIHSEFEKQ 344  
 DB 237 ITTKRPNNGISPMEWYNLGLKIAEODFIPDELIHSEFEKQ 337

RESULT 5  
 ID Q9ALM6 PRELIMINARY; PRT; 339 AA.  
 AC Q9ALM6;  
 DT 01-JUN-2003 (TREMblrel. 17, Created)  
 DT 01-JUN-2003 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)

DE Putative N-acetyl neuraminic acid synthetase Neub.  
 GN NEUB.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 NCBI\_TaxID=1311;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NT6;  
 RA McKinnon K., Chaffin D.O., Rubens C.E.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF357958; AAK11670.1; -  
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro; IPR006014; Antifreeze dom.  
 DR InterPro; IPR006190; Antifreeze \_like.  
 DR InterPro; IPR004144; Neug.  
 DR Pfam; PF01354; Antifreeze; 1.  
 DR Pfam; PF03102; Neug; 1.  
 DR PROSITE; PS50844; AFP\_LIKE; 1.  
 KW SEQUENCE 339 AA; 37773 MM; 156C2DBEFD453C7 CRC64;  
 SQ

Query Match 55.1%; Score 979.5; DB 2; Length 339;  
 Best Local Similarity 56.3%; Pred. No. 4.4e-61;  
 Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

QY 4 IYVAEIGCNHNSVDIAREMILKAEAGVNAVKQTFKADKLISALAPRAEYQKNTGE 63  
 DB 2 VYIIAEIGCNHNSVDIAREMILKAEAGVNAVKQTFKADKLISALAPRAEYQKNTGT 61  
 QY 64 LESQLEMTKLEMKYDYVILHMEYAVSLNDVSTPDESIDPLASIKOKIWKIPSGEL 123  
 DB 62 ADSQLEMTKLEMKYDYVILHMEYAVSLNDVSTPDESIDPLASIKOKIWKIPSGEI 121  
 QY 124 LNEPYLEKIAKLPIDPKKIIISTGMATIDEIKOSVSIINNKPVPNGNTIIHCNTEYPTP 183  
 DB 122 TNLPLEKIKG--QOKKVILSTGMAVMEIHOAVNIIIRONGTT--DISILHCTTEYPTP 176  
 QY 184 FEDVNLAINDIKKEFPKNNIGFSDHSSGFYAALAAVEYGTTFEKFHTLDSKSGPDHL 243  
 DB 177 YPSLNLAVIHTLKDFKDLITIGSDHSIGSEVPAAAGAEVIEKHTLDTNNMGDPDK 236  
 QY 244 ASIEDDELKHLCTIGRCYKESKSGNSKYVTASERKNTIVAKSIATKIEIKGAVPSKKN 303  
 DB 237 ASATPDIIAALVKGRIVEQALGRREKIPDPVEEKNTIVAKSVVALPKIKGDIYSTEN 296  
 QY 304 ITTKRPNNGISPMEWYNLGLKIAEODFIPDELIHSEFEKQ 344  
 DB 297 ITTKRPNNGISPMEWYNLGLKIAEODFIPDELIHSEFEKQ 337

RESULT 6  
 ID Q93T12 PRELIMINARY; PRT; 339 AA.  
 AC Q93T12;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE Putative N-acetyl neuraminic acid synthetase Neug.  
 NEUB.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 NCBI\_TaxID=1311;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CNHC 1/82;  
 RA McKinnon K., Chaffin D.O., Rubens C.E.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF355776; AAK43615.1; -  
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
 DR InterPro; IPR006014; Antifreeze dom.  
 DR InterPro; IPR006190; Antifreeze \_like.  
 DR InterPro; IPR004144; Neug.

DR Pfam; PF01354; Antifreeze; 1.  
DR Pfam; PF0102; NeutB; 1.  
DR PROSITE; PS0844; AFP\_LIKE; 1.  
SQ SEQUENCE 339 AA; 37791 MW; 156C2B2BF3157C7 CRC64;

Query Match	55.1%;	Score 979.5;	DB 2;	Length 339;
Best Local Similarity	56.3%;	Pred. No. 4.4e-67;		
Matches 192;	Conservative 54;	Mismatches 90;	Indels 5;	Gaps 2

Qy	4	IYVAIEGCHNHSVJIAEEMIKKAEACVNVAFKQTKADKLISALPKEAYOIKNTGE	63
Db	2	VYIIAIBGCHNHDINLAKKMDVAVASCVDNFKQTKRAKLISKPAKAEYOKANTGT	61
Qy	64	LESQLEMTKLEKRYDDYIHLMEYAVSLNDVSTPPEDSIDFLASLKQKIMKIPGEL	123
Db	62	ADSOLEMTKRLIESFSEBYEMEDYALSKCVZFTSPPEESIEFLISTDMPYIIPGRI	121
Qy	124	LNIPYLEKIMKLEPIPKKIIISGCAITIDEIKOSYIFINNKVCPYGNITLIHCNTAYPT	183
Db	122	TNIPYLEKIOX---QQKXILSTGMAVMEIHQVANIIRKQNTG--DISLIHCITAYPT	176
Qy	184	FEDVNLNAINDLKKHPPKNNIGFSDBSSSEFYAALNAPRYGITFIKIKHTLIDKXSGGDHL	243
Db	177	YPSNLNINVIYITLDERKDFLTIGSDHSISBEVPIAAMAGAEVIRKHTFLITNKBGDPHK	236
Qy	244	ASIEPDELKELICIGVCEVSKLSGNSKNVYTAJAEGRQKIVARKSIILAKIEIKGSEVSEK	303
Db	237	ASATPDLIALVAGVNIIVQALGRPEKIPDPVEEKIIVARKSVALKPIKGDYISIE	256
Qy	304	ITTKRPANGISPMKNTNLGKTLAEQDFITDELIIISFENQ	344
Db	297	ITVRRPANGISPMNVYDIISQEAQODFFEEDEVIROSFRFENQ	337

RESULT 7	
Q9RPCO	
ID Q9RPCO	PRELIMINARY; PRT; 341 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)  
DE CPSPN (NEUB).  
GN CPSPN OR NEUB.  
OS Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=1311;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COH1;  
RA MEDLINE=93360815; PubMed=835611;  
RX Rubens C.E., Heggen L.M., Haft R.F., Weesels M.R.;  
RT "Identification of cpsD, a gene essential for type III capsule  
RL expression in group B streptococci.";  
RL Mol. Microbiol. 8:843-855(1993).  
RX [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COH1;  
RA Chaffin D.O., Yim H.H., Beres S.B., Sweet E.S., Nitayajarn A.,  
RX Rubens C.B.;  
RT "Capsular Polysaccharide Synthesis Locus of Streptococcus agalactiae,  
RL Serotype III.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RX [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CNCTC 1/82;  
RA McKinnon K., Chaffin D.O., Rubens C.E.;  
RT "Streptococcus agalactiae type V polysaccharide synthesis operon  
RL complete sequence.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RX EMBL; AF163833; AAD53074.1; -  
DR EMBL; AF349539; AAK29661.1; -  
RX PIR; T44651; T44651.

DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA

DR InterPro; IPR006014; Antifreeze\_dom.  
DR InterPro; IPR006190; Antifreeze\_like.  
DR InterPro; IPR004144; Neub.  
DR Pfam; PF01354; Antifreeze; 1.  
DR Pfam; PF03103; Neub; 1.  
DR PROSITE; PS50844; AFP\_like; 1.  
SEQUENCE 341 AA; 3603 MW; 148B756C2B2F314 CRC64;

Query Match	55.1%	Score 979.5;	DB 2;	Length 341;
Best Local Similarity	56.3%	Pred. No. 4.4e-61;		
Matches 192;	Conservative 54;	Mismatches 90;	Indels 5;	Gaps 2.

```

Qy 4 IYVAISIGNHNHGSVDIAREMIKKKEGVNAVRFQFKDKIISALAPAEVQIKNTGE 63
Db 2 VIIHBIIGNHNGDINLAKKNVDVAVSCGDVAIVFQIFIKBEKLIKSAPAEVQAKTGT 61
Qy 64 LESQLEMTKLEMKYDDYIHLMEYAVSLNDVESTPEDEDSIDPLASLKQIKINPSEL 123
Db 62 AQSQLEMTKRLTSEFEYILBEMEDYALISGVATFTSPDESELEFLISTDMPYIKISGEI 121
Qy 124 IMLPEYETAKPIPIEDKKIISTGATIDELKQSVAFIPINKKVVGNITLHCNTEYPTP 103
Db 122 TMLPEYETKIGK---QOKKVIISTGVAWEIHOAVNLIRQGTG--DISILHCNTEYPTP 176
Qy 184 FEDVNLINAIINDLKKEHFPKKNIGCSHSHSGCFALAIANPYGITTEKHPITLDSMSGGDHL 243
Db 177 YPSLNLINAVIHTLKDDEKQDITIGYSHSISGSEVPFALAAAMGAEVIEKHTILTINNEGDOHK 236
Qy 244 ASIEPEDELKHLICIGRVCYEKSLGSNSKVVITASEROKYIARXSIIIAKTEIKKGEVFSKXN 303
Db 237 ASATPDIILAAUVKGVRIYEQALGRPEKIPDPVEEKKIYAKRSVVALKPIKKQIYISIN 256
Qy 304 ITTKRPNGISIPMEWNLIGKIAEDDFIDELIHSERKN 344
Db 297 ITVKRPNGISIPANVDILIGQERQDDFEDEVIDRSEFANO 337

```

RESULT 8	
Q8DZ3	
ID Q8DZ3	PRELIMINARY; PRT; 341 AA

DR 01-MAR-2003 (TEMBrel, 23, Created)  
DT 01-MAR-2003 (TEMBrel, 23, last sequence update)  
DT 01-JUN-2003 (TEMBrel, 24, last annotation update)  
DE N-acetyl neuraminic acid synthetase NeuB.  
GN NEUB OR SAG1161.  
OS Streptococcus agalactiae (serotype V).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=216466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2603 V/R / serotype V;  
RX MEDLINE=22222988; PubMed=122005447;  
RA Tetelein H., Maingram V., Czelusniak M.J., Eisen J.A., Peterson S.,  
RA Wessels L.C., Paulsen I.T., Neilson K.E., Margalit I., Read T.D.,  
RA Masloff L.C., Wolf A.M., Bean M.J., Brinkac L.M., Daugherty S.C.,  
RA Masoff R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
RA Radue D., Fedocova N.B., Scanlan D., Khouri H., Mulligan S., Mora M.,  
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarsell M., Mora M.,  
RA Jacobini E.T., Bretton C., Galili G., Mariani M., Vengli F., Malone D.,  
RA Rinaudo D., Rapinotti R., Telford J.L., Kasper D.L., Grandi G.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative genomic analysis of an  
RT emerging human pathogen, serotype V *Streptococcus agalactiae*,"  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
RL EMBL; AE014245; AN000043.1; -.  
DR PIR; T44651; T44651.  
DR TIGR; SAG1161; -.  
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro; IPR006014; Antifreeze dom.  
DR InterPro; IPR006190; Antifreeze-like.

DR InterPro: IPR004144; Neut.  
DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF01302; Neut; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
KM Complete proteome.  
SQ SEQUENCE 341 AA; 38033 MW; 148B756C2B2BF314 CRC64;

Query Match 55.1%; Score 979.5; DB 16; Length 341;  
Best Local Similarity 56.3%; Pred. No. 4.4e-61;  
Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

QY 4 IYIVAEIGCHNGSVDIREMILAKKAGVAVKQFQKADKLSALAPKAEYQIKRTGE 63  
DB 2 VYIIAEIGCHNGSINIAKKKAVDAVSCGVDAVAFQTFKAEKLSKFAKAEYQKATGT 61  
QY 64 LESQLEMTKLEMKYDYILHMEYAVSLNDVSTPDESDIDFLASLKQIKWIPSGEL 123  
DB 62 ADSQLEMTKLEMTSEFEYLEMRDYALISKVETFPSTPDESDLEFLISTDMPYIKPSGEI 121  
QY 124 LNTPLYEKIATLPIDPKKIISTGMATIDEIKQSVSIFINNKPVPVGNITLIHCNTEYPTP 183  
DB 122 TMLPYLEKIGK--QOKKVIISTGMVMEIHOAVNIIHQNGTT--DISTLHCTTEYPTP 176  
QY 184 FEDVNLAINDLKHPNNKNGFSDHSGFYAALAAVPGITTFLEKHTLIDKSMGPDHL 243  
DB 177 YPSLNLNVHITLKDSEFKDLITIGYSDHSGSEVPILAAAMGAEVIEKGFITLDTNMEGPDHK 236  
QY 244 ASIEPDELKHLCTGRCVCKSLGNSKVTASERKAKIVAKSIIATKTIKKGEVSEKN 303  
DB 237 ASATPDILALVGVKRVYEQALGRFEKIPDPVEERKNIIVAKSVVALKPKIKGDIYSIEN 296  
QY 304 IITKRPNGISPMENYNLGKIAEQDFIDELIHSSEFNQ 344  
DB 297 IITKRPNGISPMENYDLIGQEAQDDFEDEVIDRSRFEHQ 337

RESULT 9  
Q9A016 PRELIMINARY; PRT; 341 AA.  
ID Q9A016;  
AC Q9A016;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)  
DE Neut.  
GN Neut.  
OS Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=1311;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Miyake K., Matanabe M., Iijima S.;  
RT "GenJ of Streptococcus agalactiae type 1b shows beta-1,3-  
RT galactosyltransferase activity".  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB050723; BAB21605.1;  
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro: IPR006014; Antifreeze\_dom.  
DR InterPro: IPR006190; Antifreeze\_like.  
DR InterPro: IPR004144; Neut.  
DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF01302; Neut; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
SQ SEQUENCE 341 AA; 38037 MW; 300AEB4925D2F9D CRC64;

Query Match 53.5%; Score 951.5; DB 2; Length 341;  
Best Local Similarity 54.8%; Pred. No. 4.2e-59;  
Matches 187; Conservative 55; Mismatches 94; Indels 5; Gaps 2;  
QY 4 IYIVAEIGCHNGSVDIREMILAKKAGVAVKQFQKADKLSALAPKAEYQIKRTGE 63  
DB 2 VYIIAEIGCHNGSINIAKKKAVDAVSCGVDAVAFQTFKAEKLSKFAKAEYQKATGT 61

QY 64 LESQLEMTKLEMKYDYILHMEYAVSLNDVSTPDESDIDFLASLKQIKWIPSGEL 123  
DB 62 ADSQLEMTKLEMTSEFEYLEMRDYALISKVETFPSTPDESDLEFLISTDMPYIKPSGEI 121  
QY 124 LNTPLYEKIATLPIDPKKIISTGMATIDEIKQSVSIFINNKPVPVGNITLIHCNTEYPTP 183  
DB 122 TMLPYLEKIGK--QOKKVIISTGMVMEIHOAVNIIHQNGTT--DISTLHCTTEYPTP 176  
QY 184 FEDVNLAINDLKHPNNKNGFSDHSGFYAALAAVPGITTFLEKHTLIDKSMGPDHL 243  
DB 177 YPSLNLNVHITLKDSEFKDLITIGYSDHSGSEVPILAAAMGAEVIEKGFITLDTNMEGPDHK 236  
QY 244 ASIEPDELKHLCTGRCVCKSLGNSKVTASERKAKIVAKSIIATKTIKKGEVSEKN 303  
DB 237 ASATPDILALVGVKRVYEQALGRFEKIPDPVEERKNIIVAKSVVALKPKIKGDIYSIEN 296  
QY 304 IITKRPNGISPMENYNLGKIAEQDFIDELIHSSEFNQ 344  
DB 297 IITKRPNGISPMENYDLIGQEAQDDFEDEVIDRSRFEHQ 337

RESULT 10  
Q9F508 PRELIMINARY; PRT; 344 AA.  
ID Q9F508;  
AC Q9F508;  
DT 01-MAR-2003 (TREMREL. 23, Created)  
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)  
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)  
DE Putative N-acetyl neuraminic acid synthetase Neut (EC 4.1.3.-).  
GN NAB1 OR LA1609.  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OC NCBI\_TaxID=173;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB011338; AAN48808.1;  
DR GO; GO:0016829; P:lyase activity; IEA.  
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR InterPro: IPR006014; Antifreeze\_dom.  
DR InterPro: IPR006190; Antifreeze\_like.  
DR InterPro: IPR004144; Neut.  
DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF01302; Neut; 1.  
DR PROSITE: PS50844; AFP\_LIKE; 1.  
KM lyase; Complete proteome.  
SQ SEQUENCE 344 AA; 38200 MW; 985DFOE0340F1792 CRC64;

Query Match 48.8%; Score 868; DB 16; Length 344;  
Best Local Similarity 52.7%; Pred. No. 3.2e-53;  
Matches 116; Conservative 54; Mismatches 100; Indels 4; Gaps 2;  
QY 3 NIYIVAEIGCHNGSVDIREMILAKKAGVAVKQFQKADKLSALAPKAEYQIKRTGE 62  
DB 13 NIIYIIEGCHNGSINIAKKKAVDAVSCGVDAVAFQTFKAEKLSKFAKAEYQKATGT 72  
QY 63 ELESQLEMTKLEMKYDYILHMEYAVSLNDVSTPDESDIDFLASLKQIKWIPSGEL 122  
DB 73 EGGTQVSMKKELESVQDHFRLIEICKKNIQPLSTAFULKSIDLLIEIGLSIMKIPSGE 132  
QY 123 LNTPLYEKIATLPIDPKKIISTGMATIDEIKQSVSIFINNKPVPVGNITLIHCNTEYPTP 182  
DB 133 IINYPYLEKIGEL--NOKIILSTGMVMEIHOAVNIIHQNGTT--DISTLHCTTEYPTP 189  
QY 183 FEDVNLAINDLKHPNNKNGFSDHSGFYAALAAVPGITTFLEKHTLIDKSMGPDHL 242  
DB 190 FFEVNLVAMETLKSAF-PAKVGYSDHTVIGIVAAVAGVAVIEKGFITLDTNMEGPDH 248  
QY 243 ASIEPDELKHLCTGRCVCKSLGNSKVTASERKAKIVAKSIIATKTIKKGEVSEKN 302  
DB 249 KASLEPDELKTVSSIRNTKSNIGIKPSSSERNNIIVAKSVVALTRNISKSLFSES 308

QY 303 NITTKRPGNISPMENYINLIGKIAEDFIDELI 336  
DB 309 NITTKRPGDELSPMNMWYVIGKQARDFQIDELI 342

## RESULT 11

Q9PMX2 PRELIMINARY; PRT; 334 AA.  
AC Q9PMX2;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE N-acetylneuraminic acid synthetase (EC 4.1.3.-).  
GN NUB2 OR C01327.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NOTC 11168;  
RX MEDLINE=20150912; PubMed=10686204;  
RA Parkhill J., Wren B.W., Murgall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrall B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
DR EMBL; AL139078; CAB73754.1; -  
DR FRR; D81276; D81276.  
DR GO; GO:0004190; P:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0005524; P:ATP binding; IEA.  
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR006014; Antifreeze\_dom.  
DR InterPro; IPR001969; Antifreeze\_1like.  
DR InterPro; IPR005479; Asprotease\_AS.  
DR InterPro; IPR004144; CPase\_L\_D2.  
DR InterPro; IPR004144; Neut.  
DR Pfam; PF01354; Antifreeze; 1.  
DR Pfam; PF01352; Neut; 1.  
DR PROSITE; PS00844; AFP\_LIKE; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS00867; CPASASE\_2; 1.  
KM Complete proteome.  
SQ SEQUENCE 334 AA; 37091 MW; 59FA32E8EA92CFD CRC64;

Query Match 48.8%; Score 867; DB 16; Length 334;

Best Local Similarity 51.5%; Pred. No. 3.7e-53; Mismatches 97; Indels 4; Gaps 2;

Matches 173; Conservative 62; Mismatches 97; Indels 4; Gaps 2;

QY 1 MSNIYVAEIGCHNHSVDIAREMILKAEAGVNAVKEQTFKADKLISAIAPAEYQIKN 60  
DB 1 MKKTLILAEAGVNHNDLNAKLILHAKKCNIAFSTPDESVDLNEGKFKIPS 60  
QY 61 TGLBSQLMWTKEKEMKYDDYLMMEYAVSLNDVSTPDESIDFLASLKQKIWKIPS 120  
DB 61 TANDESQLOMVOQLIEDLNAKLILHAKKCNIAFSTPDESVDLNEGKFKIPS 120  
QY 121 GELNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
DB 121 GELTNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
QY 121 GELTNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
DB 121 GELTNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
QY 181 PTPEDVNLAINLIDKHPKPNKNGISDSSGFFAALAAVPGITTEHFTLDSKSGSP 240  
DB 181 PTPEDVNLAINLIDKHPKPNKNGISDSSGFFAALAAVPGITTEHFTLDSKSGSP 240  
QY 178 PAPFNVNKKAMQSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEKHFTLDSKSGSP 236  
DB 178 PAPFNVNKKAMQSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEKHFTLDSKSGSP 236  
QY 241 DHLASIEPDELKHLICIGVRCVSKLSGNSKYVTASERKNKIIVARSKIIAKTEIKKEGVFS 300  
DB 241 DHLASIEPDELKHLICIGVRCVSKLSGNSKYVTASERKNKIIVARSKIIAKTEIKKEGVFS 300  
QY 237 DHLASIEPDELKHLICIGVRCVSKLSGNSKYVTASERKNKIIVARSKIIAKTEIKKEGVFS 296

QY 301 ENNITTKRPGNISPMENYINLIGKIAEDFIDELI 336  
DB 297 ENNITTKRPGNISPMENYINLIGKIAEDFIDELI 332

## RESULT 12

Q7X523 PRELIMINARY; PRT; 334 AA.  
AC Q7X523;  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE PtmC.  
GN PtmC.  
OS Campylobacter coli.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=195;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC167;  
RX MEDLINE=96423180; PubMed=8825781;  
RA Guerry P., Doig P., Alm R.A., Burr D.H., Kinsella N., Trust T.J.;  
RT "Identification and characterization of genes required for post-  
translational modification of Campylobacter coli VC167 flagellin.";  
RL Mol. Microbiol. 19:369-378(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC167;  
RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Guerry P.;  
RT "Structural heterogeneity of carbohydrate modifications affects  
serospecificity of Campylobacter flagellins.";  
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC167;  
RA Guerry P.M., Doig P., Alm R.A., Burr D.H., Kinsella N., Trust T.J.;  
RT Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC167;  
RA Guerry P., Ewing C.P., Moran A.P., Trust T.J.;  
RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC167;  
RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Guerry P.;  
RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY102621; NAM76227.1; -  
SQ SEQUENCE 334 AA; 36905 MW; C415F6473D91D186 CRC64;

Query Match 47.5%; Score 844; DB 2; Length 334;

Best Local Similarity 49.7%; Pred. No. 1.5e-51; Mismatches 98; Indels 4; Gaps 2;

Matches 167; Conservative 67; Mismatches 98; Indels 4; Gaps 2;

QY 1 MSNIYVAEIGCHNHSVDIAREMILKAEAGVNAVKEQTFKADKLISAIAPAEYQIKN 60  
DB 1 MKKTLILAEAGVNHNDLNAKLILHAKKCNIAFSTPDESVDLNEGKFKIPS 60  
QY 61 TGLBSQLMWTKEKEMKYDDYLMMEYAVSLNDVSTPDESIDFLASLKQKIWKIPS 120  
DB 61 TANDESQLOMVOQLIEDLNAKLILHAKKCNIAFSTPDESVDLNEGKFKIPS 120  
QY 121 GELNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
DB 121 GELTNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
QY 121 GELTNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
DB 121 GELTNLPYLEKIAKLPIPDKKIISTGMATIDEIKOSVIFINNKVPVGNITILHCNTEY 180  
QY 181 PTPEDVNLAINLIDKHPKPNKNGISDSSGFFAALAAVPGITTEHFTLDSKSGSP 240  
DB 181 PTPEDVNLAINLIDKHPKPNKNGISDSSGFFAALAAVPGITTEHFTLDSKSGSP 240  
QY 178 PAPFNVNKKAMQSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEKHFTLDSKSGSP 236  
DB 178 PAPFNVNKKAMQSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEKHFTLDSKSGSP 236  
QY 241 DHLASIEPDELKHLICIGVRCVSKLSGNSKYVTASERKNKIIVARSKIIAKTEIKKEGVFS 300

Db 237 DHKASLEPDELQJCTKIRELESALGDGIRKQASKEKRNINIAKSLVAKKIKGSIIFS 296  
 QY 301 EXNITTKRPNGISPMENYVNLGKIAEODPIPDELI 336  
 Db 297 EENLTKRPAGCISAMRYDEVLGKKASNDVEDELI 332

## RESULT 13

QY0911 PRELIMINARY; PRT; 333 AA.  
 ID QY0911  
 AC QY0911  
 DT 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Putative N-acetylneuraminic acid synthetase (BC 4.1.3.-).  
 GN NEUB, SPSE OR SYNW0448.  
 OS Synecchococcus sp. (strain WH8102).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.  
 NC NCB1\_Taxid=84588;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=22825697; PubMed=12917641;  
 RA Patenik B., Brahamsa B., Larimer F.W., Land M., Hauser L., Chain P.,  
 RA Lamerdin J., Regala W., Allen E.B., McCarten J., Paulsen I.,  
 RA Durene A., Patensky F., Webb E.A., Waterbury J.,  
 RT "The genome of a motile marine Synecchococcus".  
 RL Nature 424:1037-1042 (2003).  
 DR EMBL, BX569690; CAB06963.1;  
 KW Lyase; Complete proteome.  
 SQ SEQUENCE 333 AA; 36219 MW; 2859DC17763C94E CRC64;  
 Query Match 43.3%; Score 769; DB 16; Length 333;  
 Best Local Similarity 45.5%; Pred. No. 2.9e-46;  
 Matches 153; Conservative 62; Mismatches 117; Indels 4; Gaps 2;

OX NCB1\_Taxid=672;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.B.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6".  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBD databases.  
 DR EMBL: AB016799; A000912.1;  
 DR GO: GO:0016051; P:cardiocyte biosynthesis; IEA.  
 DR InterPro: IPR006014; Antifreeze\_dom.  
 DR InterPro: IPR006190; Antifreeze\_like.  
 DR InterPro: IPR004144; Neub.  
 DR Pfam: PF01354; Antifreeze; 1.  
 DR Pfam: PF03102; Neub; 1.  
 DR PROSITE: PSS0844; AFP\_LIKE; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 333 AA; 36504 MW; 419D1FA07EB32A4 CRC64;  
 Query Match 40.7%; Score 723.5; DB 16; Length 333;  
 Best Local Similarity 44.2%; Pred. No. 4.7e-43;  
 Matches 149; Conservative 66; Mismatches 117; Indels 5; Gaps 3;

QY 1 MSN-ITVAEIGCNHNSVDIAEMILKAEAGVNAVKPOTFKADKLISALPAKAYOIK 59  
 Db 1 MTNKTFLIAGVNHGDIILAKQLIDAAADAGVAVKEFTWKTLLVTEADAKAYOVE 60  
 QY 60 NTGELESQLEMTKKLMKYDDYLHMEYAVSLNDVSPFPDESDIDFLASLKOKIKIP 119  
 Db 61 NTREETQFEMLRKLSIDSDPTELKSYCDCKKITMSTIDESQATFLDGL-QANPKIG 119  
 QY 120 SGELNLPYLEKIAKLPIPDKKIISTGATIDEIKQSVSIFINNKPVGNIITLACNTE 179  
 Db 120 SGELTMTPLRIHASPAA--KPVILSTGMGYLSEVEHAVALTLDAGSLDMITVLATTD 176  
 QY 180 YPTPEEDVNLAINDLKHPKRNIGFSDHSSGFYAAIAVPGITFIEHFTLDSMSG 239  
 Db 177 YTPAPEDVNLAKTIEQAFPGIIVGSDHTGTEIPVAVALGAVTEGHFTLDSMSG 236  
 QY 240 PDHLASIEPDELKHLCTGVRCVKSLSGNSKVVTASERKNKIYAKRSIAKTEIKGEVF 299  
 Db 237 PDHKSLEPDELADMTATIRNTEQALNGMKVPTKTEQENRNIRVSIVAKRIMAGSII 296  
 QY 300 SEKNITTKRPNGISPMENYVNLGKIAEODPIPDELI 336  
 Db 297 CADMLEIKRPNGISIPTRWDEVGSIKDDYOIGELI 333

## RESULT 15

QY0770 PRELIMINARY; PRT; 357 AA.  
 ID QY0770  
 AC QY0770  
 DT 01-JUN-2003 (Tremblrel. 24, Created)  
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Putative N-acetylneuraminic acid synthetase.  
 GN VP0290.  
 OS *Vibrio parahaemolyticus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 NC NCB1\_Taxid=670;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=RIMD 2210633 / Serotype O3:k6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tegomori K.,  
 RA Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shingawa H., Hattori M., Iida T.;  
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism  
 RT distinct from that of *V. cholerae*.";  
 RL Lancet 361:743-749 (2003).  
 DR EMBL: AP005073; BAC58463.1;  
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.

Thu May 13 11:53:14 2004

us-09-930-440b-8.rsp

Page 8

DR InterPro: IPR006014; Antifreeze\_dom.  
DR InterPro: IPR006190; Antifreeze\_like.  
DR InterPro: IPR004144; Neut.  
DR Pfam: PF01354; Antifreeze; 1.  
DR Pfam: PF0102; Neut; 1.  
DR PROSITE: PS00844; AFP\_like; 1.  
KM Complete proteome.  
SQ SEQUENCE 357 AA; 38639 MW; 8B887C5568BC8F54 CRC64;

Query Match 39.8%; Score 708.5; DB 16; Length 357;  
Best Local Similarity 43.1%; Pred. No. 5.9e-42;

Matches 154; Conservative 60; Mismatches 114; Indels 29; Gaps 4;

QY 6 IVAELICNNGSVDIRREMLKAKAGVNAVKQTEKADLISAIAPKAEYQKNTGELLE 65  
DB 4 IIEAGVNHNGQELAPALVDAAHEGADIVKQTERKANLVTEAKQAEYQVNTKQEP 63  
QY 66 SOLEMTKLEMKYDDYLHMEYAVSNLDVSTPFQEDSIDFLAS-LKQIKWIPSGELL 124  
DB 64 SOLAMLSRELSEYEHHDLVKYNCSIGIEFLSTAFDLESIDFLVNDLGLTRLPSEGLT 123  
QY 125 NLEPYLKIATLPFOKKIISTGMATIDEIKQSVIF----- 161  
DB 124 NAFLVAHAARTGC--DLIVSTGVALTSELEALGVIAFGYTAADQEKPMLGFOEAYAS 180  
QY 162 -INNKVPVGNITLHCTEYPTPEEDVNLNAINDLKHPKONIGFSDHSSGFYAAIAAV 220  
DB 181 EAGQKALKEKVTILHCTEYPAFMAEINLMSKDTLGSADFQ-LAAGYSDHSEGITPIAAV 239  
QY 221 PYGITTEKFTLDKSNSGPDHLASIHPPDELKHLCLGVRCVEKSLGNSKNTVTAERKAK 280  
DB 240 ARGAVLIEKFTLDKMGEGPDHKSLEPQELGAVNAIRQIEVALSSVKTPESEYKAK 299  
QY 281 IVAKSIIAKTEIKGEVFPSEKNTTGRPGNGISPMEMVNLGKIAEODFIPELII 337  
DB 300 AVAKSLVAATDIKGESELTENLITKRPQSGLSPTVYMWELIGSKATKGYKAGELL 356

Search completed: May 6, 2004, 09:07:37  
Job time : 36.1171 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: May 6, 2004, 08:59:54 / Search time 14.4061 Seconds  
(without alignments)  
1239.929 Million cell updates/sec

Title: US-09-930-440b-8  
Perfect score: 1778  
Sequence: 1 MSNIIYVAIGCNHNGSVDI.....EODFIDELIHSEKNGGE 346

Scoring table: BIOSIM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents AA:\*

- 1: /cg2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cg2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cg2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cg2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cg2\_6/ptodata/2/1aa/PTCTUS.COMB.pep:\*
- 6: /cg2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	510.5	28.7	359	4	US-09-516-143A-4
2	475.5	26.7	346	4	US-09-495-406-21
3	110.5	6.2	1365	4	US-09-376-330-18
4	109.5	6.2	739	4	US-09-543-681A-6437
5	108.5	6.1	1365	6	5194600-4
6	101	5.7	772	4	US-09-543-681A-6467
7	99.5	5.6	10182	4	US-09-134-001C-3159
8	98.5	5.5	373	4	US-09-724-864-43
9	98	5.5	256	4	US-09-113-750A-35
10	96.5	5.4	525	3	US-09-113-750A-35
11	94	5.3	491	4	US-09-914-259-18
12	93	5.2	815	4	US-09-543-681A-6286
13	93	5.2	1155	4	US-08-480-604A-28
14	92.5	5.2	1296	1	US-08-405-496A-28
15	92.5	5.2	1296	3	US-08-915-136-28
16	92.5	5.2	1296	3	US-09-084-517-28
17	92.5	5.2	1296	3	US-08-961-083-106
18	92	5.2	626	3	US-09-536-784-106
19	92	5.2	626	4	US-09-789-657-15
20	92	5.2	769	4	US-08-209-521-15
21	92	5.2	770	3	US-08-961-810-123
22	92	5.2	770	3	US-08-352-902D-123
23	92	5.2	770	3	US-09-265-503B-123
24	92	5.2	770	3	US-08-809-326A-6
25	91.5	5.1	571	3	US-09-689-914A-6
26	91.5	5.1	571	4	US-09-689-913A-6
27	91.5	5.1	571	4	US-09-689-913A-6

28	91.5	5.1	571	4	US-09-689-916A-6	Sequence 6, Appli
29	91.5	5.1	580	4	US-09-198-452A-332	Sequence 32, App
30	91	5.1	461	4	US-09-345-473E-35	Sequence 35, Appl
31	90.5	5.1	760	4	US-09-107-532A-5490	Sequence 5490, Ap
32	90	5.1	458	4	US-09-158-452A-209	Sequence 209, App
33	89.5	5.0	693	3	US-08-235-636C-72	Sequence 72, Appl
34	89	5.0	566	4	US-09-134-001C-3431	Sequence 3431, Ap
35	89	5.0	580	4	US-08-913-159-10	Sequence 10, Appl
36	89	5.0	751	4	US-09-252-991A-33073	Sequence 33073, A
37	89	5.0	1087	4	US-09-914-259-12	Sequence 12, Appl
38	88.5	5.0	854	4	US-09-254-352B-18	Sequence 18, Appl
39	88.5	5.0	992	4	US-09-206-942-61	Sequence 59, Appl
40	88.5	5.0	998	4	US-09-206-942-59	Sequence 59, Appl
41	87.5	4.9	563	4	US-09-107-532A-4741	Sequence 4741, Ap
42	87	4.9	392	1	US-07-768-286B-4	Sequence 4, Appli
43	87	4.9	406	4	US-09-252-991A-22251	Sequence 22251, A
44	87	4.9	477	2	US-08-432-016-3	Sequence 3, Appli
45	87	4.9	477	2	US-08-684-594-3	Sequence 3, Appli

## ALIGNMENTS

RESULT 1	
US-09-516-143A-4	
Sequence 4, Application US/09516143A	
Patent No. 6333182	
GENERAL INFORMATION:	
APPLICANT: Human Genome Sciences, Inc.	
TITLE OF INVENTION: Human Glycosylation Enzymes	
FILE REFERENCE: PEP505PCT	
CURRENT APPLICATION NUMBER: US/09/516,143A	
CURRENT FILING DATE: 2000-03-01	
PRIOR APPLICATION NUMBER: 60/122,409	
PRIOR FILING DATE: 1999-03-02	
NUMBER OF SEQ ID NOS: 6	
SOFTWARE: Patent In Ver. 2.1	
SEQ ID NO 4	
LENGTH: 359	
TYPE: PRT	
ORGANISM: Homo sapiens	
US-09-516-143A-4	
Query Match	28.7%; Score 510.5; DB 4; Length 359;
Best Local Similarity	36.3%; Pred. No. 2,36-46;
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6	
QY	5 YIYVAIGCNHNGSVDIAREMILKAKERGVNAKFTQTRADKLISAIAPKAEYQIKNGEL 64
DB	20 FIIAEIGQNHQGDIDVAKRMIRMAKCEQADCAKQSKELRFRKRLERYTSKHSWG 78
QY	65 ESQLEMTKKLEMKYDDYHLKRYAVSLDVFSTPDEDSIDPLASTLKOKIMKIPSGEL 124
DB	79 KTYGHEKHLFSDQRELRVABEVGIFFTAGMDMAVEFLHENVFFKVGSGDTN 138
QY	125 NLPYLEKTIAPLPIDKKIISTGMATIDEIKOSVIFINNTVPVGNITLHCNTERYPPF 184
DB	139 NFPYLEKTIAPLPIDKKIISTGMATIDEIKOSVIFINNTVPVGNITLHCNTERYPPF 192
QY	185 EDVNLNAINDKKHPKKNIGFSDHSGGFVAIAAVPGIFIEKFTLDKSMGPDHIA 244
DB	193 EDVNLNAINDKKHPKKNIGFSDHSGGFVAIAAVPGIFIEKFTLDKSMGPDHIA 252
QY	245 SIEPDELHCLIGVACVKSJGNSKVYTAERKRIIVAKRSIIAKTEIKGEVFSKNT 304
DB	253 SIEPDELHCLIGVACVKSJGNSKVYTAERKRIIVAKRSIIAKTEIKGEVFSKNT 312
QY	305 TTK-RPENGISPKMWTYLLGK-----IAEDDFIDELI 336
DB	313 TVKVGEP-KAVPPEIDFNLVKKVLTVEEDDTIWERLV 350
RESULT 2	

Thu May 13 11:53:13 2004

us-09-930-440b-8.rat

Page 2

```
US-09-495-406-21
; Sequence 21, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Makarchuk, Warren W
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 21
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: static acid synthase from C. jejuni OH4384 (ORF 8a
US-09-495-406-21
Query Match
Best Local Similarity 26.7%; Score 475.5; DB 4; Length 346;
Best Local Similarity 36.7%; Pred. No. 1.3e-42;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
QY 6 IVAETGCHNGSVIAREMILKAEAGVNAVKPQTFADGLISALPKAEYQIKATGELE 65
DB 19 VPEFQIHNHNSLEIAKIMVDAFSTGAKIKQTHIVEDEMSKAKV---IPGNAKI- 74
QY 66 SLEMTKLEMKYDDYHMEYAVSLNDVFTPEDESDIDPLASLKQKIMTIPSGELN 125
DB 75 STYEMOCALDYKDELAKETKELGLVLTSPFSAGANRLKEDMGVSAFISGECKN 134
QY 126 LPYLEKIALPIPDKKIISTGMATIDIKOSVIFINRKVPVGNITILHNTETYPPE 185
DB 135 YPLIKHIAF---KKPMIVSTGNNSISIKETVKILIDNEIP---FVLMHTNLTYPPIPN 188
QY 186 DVNLVAINDLKKEHFKPNNGISGDSGFAIAAVPGYGFIEFKHTLIDKSNQGDPLAS 245
DB 189 LVPLNAMLKKER-SCWVGSDHTDNLALGVALGCVLEKHTDSMHSRSGFDVCS 247
QY 246 IEPDELKHCIVNCEKSLGSN-SKYVTSERKNTVAKSIIATKIKGEVFSEKNI 304
DB 248 MOTQALKEIITQSCQKALMGNNESKRAAKQEVITIDFAFASVSIKDIKGEVLSMNI 307
QY 305 TTKRFG-NGISPMEMYNLLGKIAEQDPIPELIIHSEF 341
DB 308 WKKRGIGISAAFEENILGKKALRDIENTQUSYEDF 345
RESULT 3
US-09-376-330-18
; Sequence 18, Application US/09376330
; Patent No. 6393321
; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignart, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGGT
; FILE REFERENCE: 2139-9US
; CURRENT APPLICATION NUMBER: US/09/376,330
; PRIOR FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 1365
```

```
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: S. cerevisiae KRES
US-09-376-330-18
Query Match
Best Local Similarity 6.2%; Score 110.5; DB 4; Length 1365;
Best Local Similarity 21.5%; Pred. No. 0.03; Indels 97; Gaps 19;
Matches 78; Conservative 60; Mismatches 127; Indels 113; Gaps 14;
QY 62 GHEESGLEMTK--KLEMKYDDYHMEYAVSLNDVSTPEDESDIDPLASLKQKIMTIP 119
DB 41 GVDVNDL-LTNLYPVLGDLDEIDQENLVALTSNVLREYRDEKADVADLELVASY--P 97
QY 120 SGEL-----LNLPLEKIALPIPD-----KKIISTGMATIDE 153
DB 98 MGVHIGDISNAEGDANSSTFVLANGRYK-----PDVFIKSKDLITQKVPDADV 151
QY 154 IK-QSVSIFINRKVPVGNITILHNTETYPPEDEVNLN---AIN-DLKHPKRNIGS 207
DB 152 ICPVDVIGTNSBAP--LIIYGPVIDSDFEFENLPMEMANGKRFIMRSTCSL 209
QY 208 DHSRGY-----AIIAAPYGIPTIEK-HFTIDKSN---SGPDHAIIEPD 249
DB 210 DKSVEYPLTHPLEITLQNGSRMSSIPQ---LKKILYVPKEILVQANDDQDLHLEPE 265
QY 250 ELKHLICIVRCEKSLGNSKVYVTSERKNTVAR-----KSIATKIKGEVFSEKNI 303
DB 266 EURELDLRTSLISFEYQKQDTATLNFYSIVNFPPLISKQIKVSVNKOITISNEB 325
QY 304 IYTK-----RPNGISPMEM-----YNLLG-----KIAS-QDPIPELIIHS 339
DB 326 LNSKGFVYMLGLYINGQNMKITSUTPEYNTLTKTEYQSLALTYMLQELPSKCLDS 385
QY 340 EP 341
DB 386 KP 387
RESULT 4
US-09-543-681A-6437
; Sequence 6437, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BERON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6437
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6437
Query Match
Best Local Similarity 6.2%; Score 109.5; DB 4; Length 739;
Best Local Similarity 19.1%; Pred. No. 0.014; Indels 113; Gaps 14;
Matches 70; Conservative 56; Mismatches 127; Indels 113; Gaps 14;
QY 37 KQTRKADLISALAPK-AYQIKNTGSLSEGLEMT-----KLEMKY 78
DB 102 KMTTSLSDNDENTPRLAREPKKKKODHDVSAQNSLPTSSPHLEKKENKCY 161
QY 79 DDTLHMAVYSLNDVSTPEDESDIDPLASLKQKIMTIPSEGLNPLTEKIALP-- 136
DB 162 DFLPLQSLQVHIDHILINKIKDEKEDY-----YRDLQQLSLNG 201
QY 137 IPDKKIISTGMATIDRIK-----QSVSIFINN-KVP----- 167
```



Db 202 IDDLKELFKSGFNELDITKAPVDYVMEVYNQIEKNPKSEYKSTLNNCNIPDSLIQQ 261  
 QY 168 --VGNITILHCNTE-----YTFPEDEVNINAIIDLKPKHPKNNGISGDSHSG 212  
 Db 262 LILGNT---HKKVMKALBEPVFPFLNFPSPFSLSDLDLNLKSLD---GFFDSSFN 314  
 QY 213 FYAIAAAPPYGTFTFIEKFTLDKMSGPDHLASTEPD---ELKHLICIGVRCVSKLG 266  
 Db 315 YLPYISISEDKINKITSHTELEKALINKKESSESIKQIABIRGLFKELITISQSSLEBIA 374  
 QY 267 SNS---KVTASERKKKIYAKSIIKTBELKKEGVSEKNIITKRRGNGISPEWYNL 322  
 Db 375 SNSISGMSLTFPLSKTRBELTKVLRHSESSQ-KLHNEQEVTER-----SLK 422  
 QY 323 GKIAEQ 328  
 Db 423 DKVEEQ 428

RESULT 5  
 5194600-4  
 ; Patent No. 5194600  
 ; APPLICANT: BUSSEY, HOWARD; BOONE, CHARLES; SOMMER, STEVE S.;  
 ; HILL, KATHRYN; MEADEN, PHILIP  
 ; TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN  
 ; ASSEMBLY AND USE THEREOF  
 ; NUMBER OF SEQUENCES: 4  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/488,316  
 ; FILING DATE: 05-MAR-1990  
 ; SEQ ID NO: 4  
 ; LENGTH: 1365  
 5194600-4

Query Match 6.1%; Score 108.5; DB 6; Length 1365;  
 Best Local Similarity 21.3%; Pred. No. 0.049;  
 Matches 77; Conservative 61; Mismatches 127; Indels 97; Gaps 19;

QY 62 GELSSOLEMTK--KLEMKYDDYLLHMEYASINLDVSTPDEDSIDPLASKKIKIP 119  
 Db 41 GDVNNQL-LTNLYPLVTGLDEIDIDQENLVALTSNVRERYKEDVDLLELYASTY--P 97  
 QY 120 SGSL-----LNLPLYEKIAPLPID-----KKIISTGWTATIDE 153  
 Db 98 MGMTQHDSISNAEDDANSSYFVLANGNRYEK-----PDVFLKSKDLTIQKVPVDV 151  
 QY 154 IK-QSVSIFINNKVPVGNITILHCNTEYTFPEDEVNINAIIDL-----KHPKNNGIS 207  
 Db 152 IQPYDVVIGTNSBAPI--LILYGCPTVIDSDPEFNRLPHEANNGEGKFRFIRSTCSL 209  
 QY 208 DHSSGFY-----AIAAAPPYGTFTFIEK-HETLDKS--MSGPDH---LASTEPD 249  
 Db 210 DKQVEYVLTPLBLITLNGSRMSIFQ---LKKILYVKEKELVAGDNKKQJHDLBEP 265  
 QY 250 ELKHLICIGVRCVSKLGSNSKVTASERKKKIYAR-----KSLIAKTEIKGEVSEKNI 303  
 Db 266 ELREKIKRVTSLISEFYQKDIATLNTFTKSIYVNEPLSKOLIKVSSVAKCDIITSNEB 325  
 QY 304 ITTK-----RFGNGISPEWY-----YNLGS-----KTAE--QDPIPELIIHS 339  
 Db 326 LNSGQFDYNTGLYINGQNMKITSITPYNLLTALKTEYQSILKITTNLQLEPESKILDS 385  
 QY 340 EF 341  
 Db 386 KF 387

RESULT 6  
 US-09-543-681A-6467  
 ; Sequence 6467, Application US/09543681A  
 ; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; FILE REFERENCE: 2709,1002-001  
 ; CURRENT APPLICATION NUMBER: US/09/543,681A  
 ; CURRENT FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 6467  
 ; LENGTH: 772  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 US-09-543-681A-6467

Query Match 5.7%; Score 101; DB 4; Length 772;  
 Best Local Similarity 21.6%; Pred. No. 0.12;  
 Matches 57; Conservative 43; Mismatches 86; Indels 78; Gaps 11;

QY 79 DDYLLHMEYASINLDVSTPDEDSIDPLASKKIKIPSGSL-----LNLPLYEK 131  
 Db 70 EDYIAPGDY--SLSVYNTPEPLGEKTIILPKANEGKV--VPVTTIDKLKGVNVALSK 125  
 QY 132 IANLPIDPKKIIISTGWTATIDEIKQSVSIFINNKVPVGNITILHCNTEYTFPEDEVNIN 190  
 Db 126 LSDLP-EDTEI-----DNIEAVIPSRRLTFNINKLTLNASF--QVMMDK 167  
 QY 191 -----AINDLKHPKKNII--GPDHSGGFYAI-AAAPPYGTFTFIEK 230  
 Db 168 SPGEIIPAEILLDGVASAVILNVAFTKNTTDDYSSHNKPNFLNGLGVNIGPWRRTN 227  
 QY 231 FTLDKMSGPDHLASTEPDELKHLICIGVRCVSKLGSNSKVTASERKKKIYAKSIIAX 290  
 Db 228 YSYNNS--SGSDQKSSSD-----SEPSNTYVVRNIAVLYK 259

QY 291 TEIKGEVSEKNIITKRRGNGIS 314  
 Db 260 SMLKIGELITGNIIFSISPIKGIS 283

RESULT 7  
 US-09-134-001C-3159  
 ; Sequence 3159, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3159  
 ; LENGTH: 10182  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3159

Query Match 5.6%; Score 99.5; DB 4; Length 10182;  
 Best Local Similarity 17.8%; Pred. No. 13;  
 Matches 59; Conservative 71; Mismatches 115; Indels 87; Gaps 14;

QY 17 SYDIAEMKILAKEA--GVNAVRFOTFKADKLISAI--APKAEY-----QIKNGTELESQ 67  
 Db 8897 AIDTFKODIKTKDALNGISIRLTPAKSKAKELDSLFINKAQFTHANDEIMNTNSI--- 8953  
 QY 68 LEWTKKLEMKYD-----DYLLHMEYASINLDVSTPDEDSIDPLASKKIKI 115  
 Db 8954 AOLSRIVNCAFPLDANKSLHDELINQAFPVQASNYTNS--DEDLKQOPDHALSNARKV 9011

QY 116 WKIPSGELINLPYLEKIAKLPIDPKKIIISTGKATIDELIKOSVITFINNKVPVGNITIIH 175  
 DB 9012 LAKENKNDKDEKIOGL-KQVIEDTKDALN-GIORSKAKAKAIQYQ----- 9057  
 QY 176 CHTETPTPEVDVNAINDLKHPKPNIGSPDHSGFYAALA----- 218  
 DB 9058 -----SLSTINDKQRIANNNHNSDLSLANTSKASDLDNNKDLADTIE 9105  
 QY 219 ----AVPGITFI-EKHPTLD-----KSMGPHLASIEPDEIKHLCTGVRCVE 262  
 DB 9106 SNSTSVNSGVNINADKNQIEFEDEALQOASATSSKTSSENPATIE--EVLGSLQAIYDTK 9163  
 QY 263 KSLGNSKVTATSERKQKIYARSKIATKEIK 294  
 DB 9164 NALNGEORLATERKSKOLKIKGLKXDKQALB 9195

RESULT 8  
 US-09-724-864-43  
 ; Sequence 43, Application US/09724864  
 ; Patent No. 6380362  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Watson, James D.  
 ; APPLICANT: Marison, James G.  
 ; TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
 ; FILE REFERENCE: 11000,105001  
 ; CURRENT APPLICATION NUMBER: US/09/724,864  
 ; PRIOR FILING DATE: 2000-11-28  
 ; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 43  
 ; LENGTH: 373  
 ; TYPE: PRT  
 ; ORGANISM: Mouse  
 ; US-09-724-864-43

Query Match 5.5%; Score 98.5; DB 4; Length 373;  
 Best Local Similarity 21.2%; Pred. No. 0.069;  
 Matches 63; Conservative 48; Mismatches 83; Indels 103; Gaps 16;  
 QY 68 LEMTKLEMKYDYDYLHME-YAVSLNLDVFSPTPEDSDIDPLASLKOKIWKIPSGEL--- 123  
 DB 102 LQNNKIKEIKENDFKSITSLYALILNNKLTIKHPK---TFLTKRLRLYLSHNQLSRI 158  
 QY 124 -LILPLEKIAKLPIDPKKIIISTGKATIDELIKOSVITFINNKVPVGNITIIH 174  
 DB 159 PLVLP--KSLAEKRIHDKNKYKIQOTFGSMNLAHLEWNSANPLENNGJEPGAFBGVYTF 216  
 QY 175 HCTETPTPEVDVNAINDLKHPKPNIGSPDHSGFYAALAIVPGIT--TFLKHPHT 232  
 DB 217 HIRIA-----EAKTISIPKGLPFTLEHLD 242  
 QY 233 LBSMSGPHLASIEPDEIKHLCTGVRCVE-KSLGNSKVTATSERKQKIYARSKIATKEIK 291  
 DB 243 FNR-----ISTVELDLR---YRELQRLGLGNN-----KIT 271  
 QY 292 EIKKG-----EVSSEKNITTKRPGNGISPMW-----YNLLGKIAEQDPIF 332  
 DB 272 DINGTFANIPRVRBLHLENKLIKIP-SGLDELKYLQIIFLHNSIAKGVNDPCP 327

RESULT 9  
 US-09-107-532A-4072  
 ; Sequence 4072, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,532A  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arinello, Pamela Denise  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 4072:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 256 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium

FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1) LOCATION 1...256  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4072:  
 US-09-107-532A-4072

Query Match 5.5%; Score 98; DB 4; Length 256;  
 Best Local Similarity 25.7%; Pred. No. 0.042;  
 Matches 47; Conservative 26; Mismatches 76; Indels 34; Gaps 7;  
 QY 29 KEAGNAVAFQTFKADKLISAIAPAEYQIKNTGELBSQI-EMTKLEMKYDYDYLHMEY 87  
 DB 101 KQGVNKAKFVYD---VSDILQMDVXDLTQDIESQLKDRYKLS---SDQLKEVSN 153  
 QY 88 AVSLNLDVFSPTPEDSDIDPLASLKOKIWKIPSGELINLPYLEKIAKLPIDPKKIIIST 146  
 DB 154 GETKITIDQGVF-EFSVEVQANDQAVYIVNESSPEIISSEDLILITPVSDKN----- 207  
 QY 147 GKATIDELIKOSVITFINNKVPVGNITIIHCTETPTPEVDVNAINDLKHPKPNIGF 206  
 DB 208 -----GEPKQVWYI-----PSSMSQPFBEVYKIVSTGVKQNFANCMDF 248  
 QY 207 SDH 209  
 DB 249 VTH 251

RESULT 10  
 US-09-113-750A-35  
 ; Sequence 35, Application US/09113750A  
 ; Patent No. 6294176  
 ; GENERAL INFORMATION:  
 ; APPLICANT: David E. Junker and Mark D. Cochran  
 ; TITLE OF INVENTION: Recombinant Raccoonpox virus  
 ; NUMBER OF SEQUENCES: 58  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White

Thu May 13 11:53:13 2004

us-09-930-440b-8.ra1

Page 5

STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/113,750A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 55744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)262-0400  
TELEFAX: (212)664-0525  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-113-750A-35

Query Match 5.4%; Score 96.5; DB 3; Length 525;  
Best Local Similarity 23.5%; Pred. No. 0.2;  
Matches 50; Conservative 36; Mismatches 96; Indels 31; Gaps 8;

QY 12 CHNNGSVIAREMILKAEAGVNAVKFQTFRADKLISAIAPKAEYOINNGELSOLEMT 71  
DB 172 CHHSSVYRFGKCIDNIYEDNNTKIEPNF--DLSN--LIDAEVIRVTDASTQIWEK 227  
QY 72 KKLKMKDYDIHMEYANSLADVFSTPDESDIDFLASLKQKTKIPSGELNL---- 126  
DB 228 KSVLDRYVESYSRKYSKSHVYKGFSDYVRKDDIM--NIVKEL--LSNGASLTINDGSR 283  
QY 127 --PYLEKIAKLPIPKKIIISTGMATIDEIKOSVSIFINN-----KVPYGNITIL 174  
DB 284 WDPILIVRRITIMLMDIINNHTTIDKTYIHAIYINNYRNDYPRFRLVITN---K 340  
QY 175 HCNTEYPTPEFEDV--NINAINDLKHPKONI 204  
DB 341 HCLAKYTNHDDIYGTPLHMLSNKKLITPMYI 373

RESULT 11  
US-09-314-268-3  
Sequence 3, Application US/09314268  
Patent No. 6346377  
GENERAL INFORMATION:  
APPLICANT: Doobar, John  
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA  
TITLE OF INVENTION: VIRUSES  
FILE REFERENCE: 3789/80902  
CURRENT APPLICATION NUMBER: US/09/314,268  
CURRENT FILING DATE: 1999-03-19  
EARLIER APPLICATION NUMBER: 09/314,268  
EARLIER FILING DATE: 1999-05-18  
NUMBER OF SEQ ID NOS: 179  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 3  
LENGTH: 491  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-314-268-3

Query Match 5.3%; Score 94; DB 4; Length 491;  
Best Local Similarity 23.2%; Pred. No. 0.33;  
Matches 78; Conservative 38; Mismatches 128; Indels 92; Gaps 14;

QY 49 AIAPKAEYOIKNGELSOLEMTKLEMKDYDIHMEYANSLADVFST-----P 99  
DB 135 AIAKPHIITATPERLIDHLENTKGRNLAKTLVDEBARILMD--FETEVKILKIVP 193  
QY 100 FDESDIDFLASLKQKTKIPSGELNLPLYEKIAKLPIPKKIIISTGMATIDEIKOSVS 159  
DB 194 RDKRTFESADMTKKVQKQALKN-----FV---KCAVSKYQTVEXDQ--YY 239  
QY 160 IFINNKVP-----VGNITLHCNTEYPTPEFEDVNLAINDLKHPKNNIGESD 208  
DB 240 IFIPSKEDYLYIINELAGNSFMTFCST-----CNETQRTALLIRNLGP-- 285  
QY 209 HSGGFYAIAAVPYGTFPIKHEFTLDKSNKGPDLASIEPDELKHLCI-----GVR 259  
DB 286 -----TALPLHQMSK-----RLGSLNFKAKARSILATDVASRGLD 323  
QY 260 CEKSLGNSKVVYASERKKKIYAR-----KSLIAKTEIKGEVFS--EKNTTKRP 309  
DB 324 IPRDYVYVNFIDIPHSKDYIHRVGRTPRAGRSKALITFYQYDV--ELFORIEHLIGKLP 382  
QY 310 GNGSPFWEWNLGKIAE--ODFIDELIHSERKNO 344  
DB 383 GPFQDDEVNMLTERVAEORFARMELREHGEKKKK 418

RESULT 12  
US-09-914-259-18  
Sequence 18, Application US/09914259  
Patent No. 6495336  
GENERAL INFORMATION:  
APPLICANT: Makowski, Lee  
APPLICANT: Hyman, Paul  
APPLICANT: Williams, Mark  
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
FILE REFERENCE: 8471-010-999  
CURRENT APPLICATION NUMBER: US/09/914,259  
CURRENT FILING DATE: 2000-11-21  
NUMBER OF SEQ ID NOS: 180  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 815  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-09-914-259-18

Query Match 5.2%; Score 93; DB 4; Length 815;  
Best Local Similarity 19.9%; Pred. No. 0.98;  
Matches 65; Conservative 49; Mismatches 90; Indels 122; Gaps 14;

QY 83 HMEYANSLN-----LDVSTPFD--EDSIDFLASLKQKTKIPSGELNLPLYEK 131  
DB 96 HMEGVIGDGLSGIIPRIYADIFNHTYSMDENLQF--HKKVSYTEL-----YNEK 144  
QY 132 IAKLPIPKKIIISTGMATIDEIKOSVSIFINNKPVGNITILHCNTEYPTPEFEDVNLAIN 191  
DB 145 IRDLDPDEKYNL-----SIHEK-----NRVP-----YKGAITERVGGDEVLQA 185  
QY 192 INDLKHPKNNIGFSDHSGFYAIAAVYGTTFPIKHEFTLDKSNKGPDLASIER-- 246  
DB 186 IEDGKSNRVAAYTNMNEHSSRSV-----FLITVQEHQTKQKLGKYLVDLAGSEK 240  
QY 249 -----DELKHL-----CIGV-----RCVEKSLGNSK-- 270  
DB 241 VSKTGAQGYTLERAKNINKSLTALGIVISALAGTSHPVYRSKULRIQESLGNSRT 300  
QY 271 --VVTASER-----KNTIVAKSITAKTEIKGEVFSERK----- 303  
DB 301 TVIICASPSPHFNEATKSTLLFGARAKITGNVQVINEELTAEEMKGRYERKEKENTRLAA 360

Thu May 13 11:53:13 2004

us-09-930-440b-8.ra1

Page 6

QY 304 -----ITTKRPGNGISPMENYVL 321  
Db 361 LLDAAALELSRMAPGESVSEVENYVL 386

RESULT 13  
US-09-543-681A-6286  
Sequence 6286, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709, 1002-001  
CURRENT APPLICATION NUMBER: US/09/543, 681A  
PRIOR FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6286  
LENGTH: 1155  
TYPE: PR1  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6286

Query Match 5.2%; Score 93; DB 4; Length 1155;  
Best Local Similarity 20.3%; Pred. No. 1.7;  
Matches 69; Conservative 57; Mismatches 110; Indels 104; Gaps 17;

QY 62 GELSSQ--LEMTKLEMKDYHLMYAVSLNLYVSTPPEDSI-DELSLKQIK 117  
Db 85 GDLGQALLLDNITQQLKADY-----NKTVEQLPKRSIQNQLSKQKIAN 134  
QY 118 IPGGLMLPYLEKIALPIPDKIIISTGATIDEIKOSV----- 158  
Db 135 KEVD-----YQSLSLP-----LATLSQLEEVLAGLAKAGEDLANYSNELIVLT 182  
QY 159 -----SIFNNKVPVANTIT-----LHCTEYTPREDVNAINDLKHF 201  
Db 183 QPERAQSFLFNSEERLQRIALNKSADRAQSSVQLLQYLLQCSNFORLTQ 242  
QY 202 NNTGFS-----DHSSGFYALAAVYGITFTLEKFTL-----DKSMGPDHLA-SIE 247  
Db 243 SNVQLSLQLQGRDYSA-----IDLSQEHAGLQELISKRIDSSEVAKKAO 292  
QY 248 PDLKHCIGVRC-----VEKSLGNSKVTASERKATYAKSIIAKTEIKGEVS 300  
Db 293 TALSINQAIKNAFYLAQADINKL--SDKLITTONNEL--RHSIMVGNRLDRA-109 348  
QY 301 EKNTTKRPGNGISPMENYVLGKI--AEODPIDELIIH 338  
Db 349 ERHAKQ-----IDVTKGSLLSRLFEEDIEHPDOLFIN 383

RESULT 14  
US-08-480-604A-28  
Sequence 28, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPED-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-604A-28

Query Match 5.2%; Score 92.5; DB 1; Length 1296;  
Best Local Similarity 17.8%; Pred. No. 2.4;  
Matches 76; Conservative 71; Mismatches 151; Indels 129; Gaps 17;

QY 5 YVVAIGCNHNSVDIAEMKIKAYAGVNAVKQTFPADKLISALAPAEYQIKNTGEL 64  
Db 712 YVIVMLAKVNTQIDILK-----KMKEA-----L 736  
QY 65 ESQLEMTK-LEMKYDQVHLMYAVSLNLYVSTPPEDSIDPLASIKOKIKIPSGEL 123  
Db 737 ENQAEATKAIIVQNTQTEBEKNININIIDLSKINESINKMININKPLNQSIVYL 796  
QY 124 LN-LPY-LEKIAKLPIPDKII-----STGATIDEIKQSVSIFNNKVP----- 167  
Db 797 KMSMIFYGKRLIEDPDAISKDALIKYIVDNRTGLIGVDRLDKVNNLTSTDIPOLSKY 856  
QY 168 VGNITILHCTEYTPPEDVNL-----NAINDKHPKPNNG----- 205  
Db 857 VDNQSLSTFEYINNIINTSLNIRYNSNHLIDSRASKINISKVPFIDPKQIOL 916  
QY 206 FDSHSGFYAALV-APYGITFTLEKFTLDKSMGPDHLASIEPDLKRLCIGVCEVS 264  
Db 911 FVLESKIEVLKNAIVNSWY--ENFSTFWIRIPKFNIS--SLNNEYITINCMEN 971  
QY 265 LG-----SNKVTASE-----RQKIVA 283  
Db 972 SKMKVSLVGEIIMTLQDQIKQVVERYSOMINISYINKMIFVTTNNELNNSKIYI 1031  
QY 284 KSIYAKTEIKK-GEVFEKNITTRPG-----NGISPMENYVLQKIAEODPIDELIIH 338

Thu May 13 11:53:13 2004

us-09-930-440b-8.rat

Page 7

Db 1032 NGRLLDQKPSNLGNTHASNNIMFKDGCRODTHRYIWIYFNLPDEKLEKEIKD---LY 1088

QY 339 SEFKNOG 345  
Db 1089 DNQNSNG 1095

RESULT 15

US-08-405-496A-28  
Sequence 28, Application US/08405496A  
Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

TITLE OF INVENTION: NEUROTOXIN

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/405.496A

APPLICATION NUMBER: 424

FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01308

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 1296 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-405-496A-28

Query Match 5.2%; Score 92.5; DB 2; Length 1296;

Best Local Similarity 17.8%; Pred. No. 2.4;

Matches 76; Conservative 71; Mismatches 151; Indels 129; Gaps 17;

QY 5 YIVAEICGHNHGSVDIAEMLIKAEAGNAVKPOTFRKADKLSAIAKAEYQIKNGEL 64

Db 712 YIVNWLAKVTDILRK---KMKEA-----L 736

QY 65 ESQLEMTKK-LEMKYDDYLHLMAYVASINLVFSTPDEDSIDFLASIKOKIWKIPSGEL 123

Db 737 ENQATATKALINYNQYTEBEKNININIDLSKLNESINKAMINIKFLNQCSVYL 796

QY 124 LN-LPY-LEKAKIPDPKII-----STGMATIDEIKOSVSIPIKVKP----- 167

Db 797 MNSMIPYGVARLEDFDASLKDALKYIYDNRGTLIGVQRLKQKVNNTLSTDIPQLSKY 856

QY 168 VGNITILHGNTEPTPEEDVN-----NAINDLKHFKNIG----- 205

Db 857 VDNQRLSTFTTEYIKNIINTSLNRYESNHLJDSRYASKNIGSKVNFDPIDKQIQ 916

QY 206 FSDHSSGFYAALA-AVPGITPTIEKFTLDKSNAGPDHLASIEPDELKHLCTGVRCEKS 264

Db 917 FNESSKIEVILKNAIVNWSY--ENFSTSPWIRIPKFNST---SLNREYTIINCENN 971

QY 265 LG-----SNSKYVTASE-----RAKRTVA 283

Db 972 SGKRVSLNGBELIWTQDQOEIKQVVFYKQMINISDYINRWI FVLTINRNLSNKIYI 1031

QY 284 RKSIIAKTEIKK-GVESEKNITTKRPG---NGISPMWYNLLGKIAEQDFIPDELIH 338

Db 1032 NGRLLDQKPSNLGNTHASNNIMFKDGCRODTHRYIWIYFNLPDEKLEKEIKD---LY 1088

QY 339 SEFKNOG 345  
Db 1089 DNQNSNG 1095

Search completed: May 6, 2004, 09:08:40  
Job time: 16.4061 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 37.4054 Seconds  
(without alignments)  
2567.492 Million cell updates/sec

Title: US-09-930-440b-8  
Perfect score: 1778  
Sequence: 1 MSNIYVAEIGCHNGSVDI.....EODFIDELIHSEFNQGE 346

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1778	100.0	346	US-09-930-440b-8	Sequence 8, Appli
2	993.5	55.9	338	US-09-767-041-24	Sequence 24, Appli
3	516.5	29.0	359	US-10-264-237-2597	Sequence 2597, Ap
4	510.5	28.7	359	US-09-984-205-4	Sequence 4, Appli
5	510.5	28.7	359	US-09-930-440b-6	Sequence 6, Appli
6	475.5	26.7	346	US-09-816-028A-35	Sequence 35, Appli
7	475.5	26.7	346	US-10-303-161-35	Sequence 35, Appli
8	475.5	26.7	346	US-10-303-118-35	Sequence 35, Appli
9	475.5	26.7	346	US-10-303-128-35	Sequence 35, Appli
10	475.5	26.7	346	US-10-303-134-35	Sequence 35, Appli
11	475.5	26.7	346	US-10-303-162-35	Sequence 35, Appli
12	426	24.0	311	US-10-156-761-10910	Sequence 10910, A
13	402	22.6	340	US-09-882-227-490	Sequence 490, App
14	381	21.4	340	US-10-335-977-7625	Sequence 7625, Ap
15	373	21.0	314	US-10-335-977-7625	Sequence 7625, Ap

16	174.5	9.8	123	14	US-10-106-628-6926	Sequence 6926, Ap
17	114	6.4	500	12	US-10-282-122A-53768	Sequence 53768, A
18	112	6.3	335	15	US-10-369-493-8945	Sequence 8945, Ap
19	109	6.1	601	12	US-10-282-122A-55143	Sequence 55143, A
20	107	6.0	1577	15	US-10-369-493-6924	Sequence 6924, Ap
21	107	6.0	1577	15	US-10-369-493-6925	Sequence 6925, Ap
22	107	6.0	1577	15	US-10-369-493-6926	Sequence 6926, Ap
23	103	5.8	653	9	US-09-759-010-2	Sequence 2, Appli
24	103	5.8	732	14	US-10-160-748-4	Sequence 4, Appli
25	103	5.8	950	12	US-10-282-122A-7285	Sequence 47285, A
26	102.5	5.8	567	15	US-10-369-493-1053	Sequence 1053, Ap
27	102	5.7	410	12	US-10-425-114-43092	Sequence 43092, A
28	102	5.7	481	14	US-10-032-585-7372	Sequence 7372, Ap
29	102	5.7	1241	12	US-10-282-122A-11957	Sequence 51957, A
30	101.5	5.7	650	12	US-10-282-122A-74461	Sequence 74461, A
31	101.5	5.7	841	9	US-09-861-451A-30	Sequence 30, Appli
32	101	5.7	455	10	US-09-909-567B-48	Sequence 48, Appli
33	101	5.7	455	15	US-10-094-749-1646	Sequence 1646, Ap
34	101	5.7	468	12	US-10-282-122A-76865	Sequence 76865, A
35	100.5	5.7	645	12	US-10-425-114-38654	Sequence 38654, A
36	100.5	5.7	707	12	US-10-424-599-225717	Sequence 225717, A
37	100.5	5.7	707	12	US-10-424-599-225717	Sequence 21877, A
38	100	5.6	878	15	US-10-369-493-21877	Sequence 52209, A
39	99	5.6	266	12	US-10-282-122A-52209	Sequence 76807, A
40	99	5.6	381	12	US-10-282-122A-76807	Sequence 60665, A
41	98	5.6	866	12	US-10-032-585-7841	Sequence 7841, Ap
42	98	5.6	982	14	US-10-312-130-22	Sequence 22, Appli
43	98.5	5.5	358	14	US-10-312-130-22	Sequence 680, App
44	98.5	5.5	373	10	US-09-866-050A-680	Sequence 6, Appli
45	98.5	5.5	373	14	US-10-319-130-6	

## ALIGNMENTS

RESULT 1  
US-09-930-440b-8  
Sequence 8, Application US/09930440B  
Patient No. US20020142386A1  
GENERAL INFORMATION:  
APPLICANT: Beebebaugh et al.  
TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways  
FILE REFERENCE: PFS09P2  
CURRENT APPLICATION NUMBER: US/09/930,440B  
PRIOR FILING DATE: 2001-08-16  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: 60/227,579  
PRIOR FILING DATE: 2000-08-25  
PRIOR APPLICATION NUMBER: 09/516,793  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/169,624  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: 60/122,582  
PRIOR FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 8  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-930-440b-8

Query Match 100.0%; Score 1778; DB 9; Length 346;  
Best Local Similarity 100.0%; Pred. No. 4,1e+161; Indels 0; Gaps 0;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSNIYVAEIGCHNGSVDIAREMILKAEAGVNAVVEQTFKADKLISAIAPAAVQIKN 60  
DB 1 MSNIYVAEIGCHNGSVDIAREMILKAEAGVNAVVEQTFKADKLISAIAPAAVQIKN 60  
QY 61 TGBIESOLETKYKLEMKYDYDYLHMEYAVSINLDVETPEDESDIDLAISKQWIKPS 120  
DB 61 TGBIESOLETKYKLEMKYDYDYLHMEYAVSINLDVETPEDESDIDLAISKQWIKPS 120

QY 121 GELNLPYLEKIAKLPIDPKKIIISGNAITIDEIKOSVIFINNKVPVGNITILHONTY 180  
 DB 121 GELNLPYLEKIAKLPIDPKKIIISGNAITIDEIKOSVIFINNKVPVGNITILHONTY 180  
 QY 181 PTFPEVDNAINDLKKEPPKNNIGSDHSGFYAIAAVPGITFIEKFTLDKSMGSP 240  
 DB 181 PTFPEVDNAINDLKKEPPKNNIGSDHSGFYAIAAVPGITFIEKFTLDKSMGSP 240  
 QY 241 DFLASIEPDELKHLCTGVACVEKSLGNSKVTYASERKKIYAKRSIIAKTBIKGEVFS 300  
 DB 241 DFLASIEPDELKHLCTGVACVEKSLGNSKVTYASERKKIYAKRSIIAKTBIKGEVFS 300  
 QY 301 EKNTTKRPNGISPEWENLKGIAEOPFIPDELIHSEFKNGSE 346  
 DB 301 EKNTTKRPNGISPEWENLKGIAEOPFIPDELIHSEFKNGSE 346

RESULT 2  
 US-09-767-041-24  
 ; Sequence 24, Application US/09767041  
 ; Patent No. US20020055168A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Smith, Hilda  
 ; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS  
 ; FILE REFERENCE: 2183-4726  
 ; CURRENT FILING DATE: 2001-01-22  
 ; PRIOR FILING DATE: 1999-07-19  
 ; PRIOR APPLICATION NUMBER: EP98202465.5  
 ; PRIOR FILING DATE: 1998-07-22  
 ; PRIOR APPLICATION NUMBER: EP98202467.1  
 ; PRIOR FILING DATE: 1998-07-22  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 24  
 ; LENGTH: 338  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus suis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: CPSP2P  
 ; US-09-767-041-24

Query Match 55.3%; Score 993.5; DB 9; Length 338;  
 Best Local Similarity 58.4%; Pred. No. 3.5e-86;  
 Matches 199; Conservative 45; Mismatches 92; Indels 5; Gaps 2;

QY 4 IYVIAEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFKADKLISAIAPKAEYQIANTG 63  
 DB 2 VYIIAETGCHNDVDILAROMVEVAVDQGVDAVAFOTERKADLISAKPAEYQIANTG 61  
 QY 64 LESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSIDPLASLKQIKWIPSGEL 123  
 DB 62 SDQGLEMTKLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSIDPLASLKQIKWIPSGEL 121  
 QY 124 LNPYLEKIAKLPIDPKKIIISGNAITIDEIKOSVIFINNKVPVGNITILHONTY 183  
 DB 122 TNPYLEKIGR--QAQKVLISGMAVMDIHOAVKLQNGT--DISTLHCTTTEYPP 176  
 QY 184 FEDVNAINDLKKEPPKNNIGSDHSGFYAIAAVPGITFIEKFTLDKSMGSPDH 243  
 DB 177 YPALNINLVHMLTKKEFPNLTIGYSDHSGSEVPAAIAAMGAELIEKFTLDNEMEGPDH 236  
 QY 244 ASIEPDELKHLCTGVACVEKSLGNSKVTYASERKKIYAKRSIIAKTBIKGEVFS 303  
 DB 237 ASATPDLIALVGVRIEQLGFEKEPEVEYRNKIVAKRSIIAKTBIKGEVFS 296  
 QY 304 IYVIAEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFKADKLISAIAPKAEYQIANTG 344  
 DB 297 IYVIAEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFKADKLISAIAPKAEYQIANTG 337

RESULT 3  
 US-10-264-237-2597  
 ; Sequence 2597, Application US/10264237  
 ; Publication No. US2004009491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Biese et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P131P1  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450  
 ; PRIOR FILING DATE: 2000-05-19  
 ; NUMBER OF SEQ ID NOS: 2876  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 2597  
 ; LENGTH: 359  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-264-237-2597

Query Match 29.0%; Score 516.5; DB 15; Length 359;  
 Best Local Similarity 36.6%; Pred. No. 1.4e-40;  
 Matches 124; Conservative 61; Mismatches 139; Indels 15; Gaps 6;

QY 5 YVIAEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFKADKLISAIAPKAEYQIANTG 64  
 DB 20 FIIAIEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFKADKLISAIAPKAEYQIANTG 78  
 QY 65 ESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSIDPLASLKQIKWIPSGEL 124  
 DB 79 KTYGEKHLERSHQLYELQRYAEVGIPTASGMDKAEZLHELAVPEFFKSGSDTN 138  
 QY 125 MLPYLEKIAKLPIDPKKIIISGNAITIDEIKOSVIFINNKVPVGNITILHONTY 184  
 DB 139 NPYLEKIAKGR--NVISGMSQMDTKQVQIV--KPLNPFCLQCTSAVPLQ 192  
 QY 185 EDVNAINDLKKEPPKNNIGSDHSGFYAIAAVPGITFIEKFTLDKSMGSPDH 244  
 DB 193 EDVNAINDLKKEPPKNNIGSDHSGFYAIAAVPGITFIEKFTLDKSMGSPDH 252  
 QY 245 SIEPDELKHLCTGVACVEKSLGNSKVTYASERKKIYAKRSIIAKTBIKGEVFS 304  
 DB 253 SIEPDELKHLCTGVACVEKSLGNSKVTYASERKKIYAKRSIIAKTBIKGEVFS 312  
 QY 305 TTK--RPNGISPEWENLKGIAEOPFIPDELIHSEFKNGSE 346  
 DB 313 TVKVGEP--KGYPEPDIHNLVGRKVLVTEEDDTIMEELV 350

RESULT 4  
 US-09-984-205-4  
 ; Sequence 4, Application US/09984205  
 ; Patent No. US20020137175A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Coleman, Timothy A. et al.  
 ; TITLE OF INVENTION: Human Glycosylase Enzymes  
 ; FILE REFERENCE: PFS0501  
 ; CURRENT FILING DATE: 2001-10-29  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05325  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: 09/516,143  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: 60/122,409  
 ; PRIOR FILING DATE: 1999-03-02  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 359  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

Thu May 13 11:53:13 2004

US-09-930-440b-8.rapb

Page 3

US-09-984-205-4

Query Match 28.7%; Score 510.5; DB 9; Length 359;  
Best Local Similarity 36.3%; Pred. No. 5, 4e-40;  
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 5 YVAAIGCNHNSVDIAREMILAKAGVNAVVFQTFKADKLISALAPKAYQIKNTGEL 64  
DB 20 FIIAIGQNHQGDLDVAKRMIRVAKCGADCAFOKSELEFKENKALRPPYTSKSG- 78  
QY 65 ESQLEMTKLEMYDYDLHMEYASLNDVSTPPEDSIDFLASIKQIKWIPSGELL 124  
DB 79 KTYGEKRLHLEFHDQYRELQRYAEVGIPTFASGMDMAVELHLELNPFPVSGSDIN 138  
QY 125 NLPYLEKIALPLIPDKKIIISTGMATIDEIKOSVSIFINNKVFNQNTILHCTEYPTP 184  
DB 139 NFPYLEKTAKKGRP---MVISGQSWDTWKQYQIV---KPLNPFCEFLQCTSAVPLDP 192  
QY 185 EDVNLAINDLKHPKNNIGFSDHSGFYAALAAVPGYITFEKFTLTKSGSPDHLA 244  
DB 193 EDVNLAVISYQKLPDIPIGYSGHETGLAISVAALVAKVLEKHTLIDTKWQSDHSA 252  
QY 245 SIEPDELKHLICIGVRCVKSIGNSKVVYASERKNKIIVAKSIIAKTEIKGEVFSKNI 304  
DB 253 SIEPDELALVRSVRLVERALGSPTKQLPCMAKNEKLGKSVAAKVIPEGITLMDML 312  
QY 305 TTK--RPGNGISPMEMYNILGK-----IAQDFTIDELI 336  
DB 313 TVKVGEP-KAYPPEDIFNIVGKVLVTEEDDTIMEELV 350

RESULT 5

US-09-930-440b-6  
Sequence 6, Application US/09930440B  
Patent No. US2002014236A1  
GENERAL INFORMATION:  
APPLICANT: Beutenberg et al.  
TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways  
FILE REFERENCE: p5509P2  
CURRENT APPLICATION NUMBER: US/09/930,440B  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: 60/227,579  
PRIOR FILING DATE: 2000-08-25  
PRIOR APPLICATION NUMBER: 09/516,793  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/169,624  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: 60/122,582  
PRIOR FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 359  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-930-440b-6

Query Match 28.7%; Score 510.5; DB 9; Length 359;  
Best Local Similarity 36.3%; Pred. No. 5, 4e-40;  
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 5 YVAAIGCNHNSVDIAREMILAKAGVNAVVFQTFKADKLISALAPKAYQIKNTGEL 64  
DB 20 FIIAIGQNHQGDLDVAKRMIRVAKCGADCAFOKSELEFKENKALRPPYTSKSG- 78  
QY 65 ESQLEMTKLEMYDYDLHMEYASLNDVSTPPEDSIDFLASIKQIKWIPSGELL 124  
DB 79 KTYGEKRLHLEFHDQYRELQRYAEVGIPTFASGMDMAVELHLELNPFPVSGSDIN 138  
QY 125 NLPYLEKIALPLIPDKKIIISTGMATIDEIKOSVSIFINNKVFNQNTILHCTEYPTP 184  
DB 139 NFPYLEKTAKKGRP---MVISGQSWDTWKQYQIV---KPLNPFCEFLQCTSAVPLDP 192

QY 185 EDVNLAINDLKHPKNNIGFSDHSGFYAALAAVPGYITFEKFTLTKSGSPDHLA 244  
DB 193 EDVNLAVISYQKLPDIPIGYSGHETGLAISVAALVAKVLEKHTLIDTKWQSDHSA 252  
QY 245 SIEPDELKHLICIGVRCVKSIGNSKVVYASERKNKIIVAKSIIAKTEIKGEVFSKNI 304  
DB 253 SIEPDELALVRSVRLVERALGSPTKQLPCMAKNEKLGKSVAAKVIPEGITLMDML 312  
QY 305 TTK--RPGNGISPMEMYNILGK-----IAQDFTIDELI 336  
DB 313 TVKVGEP-KAYPPEDIFNIVGKVLVTEEDDTIMEELV 350

RESULT 6

US-09-816-028A-35  
Sequence 35, Application US/09816028A  
Patent No. US20020042369A1  
GENERAL INFORMATION:  
APPLICANT: Gilbert, Michel  
APPLICANT: Galtchuk, Warren W.  
TITLE OF INVENTION: National Research Council of Canada  
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
FILE REFERENCE: 019633-000111US  
CURRENT APPLICATION NUMBER: US/09/816,028A  
CURRENT FILING DATE: 2001-03-21  
PRIOR FILING DATE: US 60/118,213  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 09/495,406  
PRIOR FILING DATE: 2000-01-31  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 35  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: of sialic acid synthase from C. jejuni OH4384 (ORF 84)  
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus  
US-09-816-028A-35

Query Match 26.7%; Score 475.5; DB 9; Length 346;  
Best Local Similarity 36.7%; Pred. No. 1, 1e-36;  
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;

QY 6 YVAAIGCNHNSVDIAREMILAKAGVNAVVFQTFKADKLISALAPKAYQIKNTGEL 65  
DB 19 VPEIGINNSGLBLAKIVDAFSTGAKIIGHQTHIVEDEMSKAAKV---IPGNKI- 74  
QY 66 SLEMTKLEMYDYDLHMEYASLNDVSTPPEDSIDFLASIKQIKWIPSGELL 125  
DB 75 SIEYEMKQKADYKQELAKETKGLVYLTSPSRAGANLEBDGSAFKIGSECN 134  
QY 126 LPYLEKIALPLIPDKKIIISTGMATIDEIKOSVSIFINNKVFNQNTILHCTEYPTP 185  
DB 135 YPLIKHIAF---KPMIVSTGNSISLSIKPYKILDNELP---FVLAHTNLVYTPPN 188  
QY 186 DVNLAINDLKHPKNNIGFSDHSGFYAALAAVPGYITFEKFTLTKSGSPDHLA 245  
DB 189 LVRLNLMLEKKEF--SCWGLSDHTDNLACGAVALGACVLERHTDSMRSGPDIVCS 247  
QY 246 IEPDELKHLICIGVRCVKSIGNSKVVYASERKNKIIVAKSIIAKTEIKGEVFSKNI 304  
DB 248 KOTQALKEILIOSEQAMIRGNNSKKAQKQYITIDFAPAVSVISKDIKXGVSMDNI 307  
QY 305 TTKRPG-NGISPMEMYNILGKIAQDFTIDELIHSSEF 341  
DB 308 WYRPELGISIAAEFENILGKALRDIENDTQSYEDF 345

RESULT 7

US-10-303-161-35  
Sequence 35, Application US/10303161



Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 4

```
Publication NO. US20030148459A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,161
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRF
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: stialic acid synthase from C. jejuni OH4384 (ORF 8a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-161-35
```

```
Query Match          26.7%; Score 475.5; DB 14; Length 346;
Best Local Similarity 36.7%; Pred. No. 1,1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
```

```
QY 6 IVAEICNGNSVDIAREMILKAKAGVNAVRFQTPKADKLISALAPKAEYQIKTGELE 65
DB 19 VPEIGINHNSLEIAKIMVDAFSTGAKIKHQTHIVEDMSKAKV---IPGAKI- 74
QY 66 SOLEMTKLEMKYDYLHMEYAVSLNLDVFSTPEDESDIFLASIKQIKWIPSGELN 125
DB 75 SIYEIMQKALDYDELAKYTERKGLVYLSFSPRAGARLEDMGSAFKIGSGECNN 134
QY 126 LPELEKIAKLPIPDKKIISTGMATIDEIKOSVSIFINNKVPVGNITTLACNTEYPTPE 185
DB 135 YPLIKHIAF---KKPMIVSTGMNSIESIKPTVKILLDNEIP---FVLMHTTNLYPTPHN 188
QY 186 DVNLAINDLKKEPPKNNIGFSDHSGGFYALAAVPGYITFLKHTTLDKSMGPHLAS 245
DB 189 LVLNLMLELKEEF-SCWGLSDHTTNDNLACGAVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTIGRCVCEKSLGSN-SKVVTASERKNKIYARSKIITAKTEIKGEVPESEKI 304
DB 248 MDQALKEELIQSEQMAIMRGNESKKAQOEVTIDFAFASVVSIKDIKKEGVALSMONI 307
QY 305 TTKRPG-NGISPMWYNLIGKIAEODFIPEDLIIHSEF 341
DB 308 WKRPGLGIGISAAPENILGKKALRDIENDTQLSYEDF 345
```

```
RESULT 8
US-10-303-118-35
Sequence 35, Application US/10303118
Publication No. US20030157655A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,118
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
```

```
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRF
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: stialic acid synthase from C. jejuni OH4384 (ORF 8a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-118-35
```

```
Query Match          26.7%; Score 475.5; DB 14; Length 346;
Best Local Similarity 36.7%; Pred. No. 1,1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
```

```
QY 6 IVAEICNGNSVDIAREMILKAKAGVNAVRFQTPKADKLISALAPKAEYQIKTGELE 65
DB 19 VPEIGINHNSLEIAKIMVDAFSTGAKIKHQTHIVEDMSKAKV---IPGAKI- 74
QY 66 SOLEMTKLEMKYDYLHMEYAVSLNLDVFSTPEDESDIFLASIKQIKWIPSGELN 125
DB 75 SIYEIMQKALDYDELAKYTERKGLVYLSFSPRAGARLEDMGSAFKIGSGECNN 134
QY 126 LPELEKIAKLPIPDKKIISTGMATIDEIKOSVSIFINNKVPVGNITTLACNTEYPTPE 185
DB 135 YPLIKHIAF---KKPMIVSTGMNSIESIKPTVKILLDNEIP---FVLMHTTNLYPTPHN 188
QY 186 DVNLAINDLKKEPPKNNIGFSDHSGGFYALAAVPGYITFLKHTTLDKSMGPHLAS 245
DB 189 LVLNLMLELKEEF-SCWGLSDHTTNDNLACGAVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTIGRCVCEKSLGSN-SKVVTASERKNKIYARSKIITAKTEIKGEVPESEKI 304
DB 248 MDQALKEELIQSEQMAIMRGNESKKAQOEVTIDFAFASVVSIKDIKKEGVALSMONI 307
QY 305 TTKRPG-NGISPMWYNLIGKIAEODFIPEDLIIHSEF 341
DB 308 WKRPGLGIGISAAPENILGKKALRDIENDTQLSYEDF 345
```

```
RESULT 9
US-10-303-128-35
Sequence 35, Application US/10303128
Publication No. US20030157656A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,128
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRF
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: stialic acid synthase from C. jejuni OH4384 (ORF 8a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-128-35
```

```
Query Match          26.7%; Score 475.5; DB 14; Length 346;
```

Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 5

Best Local Similarity 36.7%; Pred. No. 1,1e-36;  
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;  
QY 6 IVAIGCHNNSVDIAREMILKAEAGVNAVKQTFPAKDLISAIAPKAEYOIKNTGELE 65  
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74  
QY 66 SLEMTKLEKDYDYLHMEYAVSLNDVSTFPEDESIDPLASLKOKIWKIPSGELIN 125  
DB 75 SIYEIMOKCALDYDELAKKEYTEKGLVYISTFSPAGANRLEDMGVSAFKISGSCNN 134  
QY 126 LPYIEKIAKLPIDPKKIISTGMAITIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185  
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188  
QY 186 DVNINAINDLKKHPKKNIGFSDHSGFYAAIAAVPYGITFEKHFTLDKSMGPDHLAS 245  
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTDNLACGVALGACVLEKHFITDSMERSGPDIVCS 247  
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVFSKNI 304  
DB 248 MDTOALKEILLIOSEQMAIMRGNNESKKAQEQVTIDPAPASVVISIKKGEVLSMDNI 307  
QY 305 TTKRPG-NGISPMWYNLLGKIAEQDFIPDELIHSEF 341  
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTQLSYEDF 345  
RESULT 10  
US-10-303-134-35  
Sequence 35, Application US/10303134  
Publication No. US20030157657A1  
GENERAL INFORMATION:  
APPLICANT: Gilbert, Michel  
APPLICANT: Wakarchuk, Warren W.  
TITLE OF INVENTION: Gangliobacter Glycosyltransferases for Biosynthesis of  
TITLE OF INVENTION: Gangliobacter Glycosyltransferases for Biosynthesis of  
FILE REFERENCE: 019633-000111US  
CURRENT FILING DATE: 2002-11-21  
PRIOR FILING DATE: 2001-03-21  
PRIOR FILING DATE: 2001-03-21  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 60/118,213  
PRIOR APPLICATION NUMBER: US 09/495,406  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 35  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a  
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
US-10-303-134-35  
Query Match 26.7%; Score 475.5; DB 14; Length 346;  
Best Local Similarity 36.7%; Pred. No. 1,1e-36;  
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;  
QY 6 IVAIGCHNNSVDIAREMILKAEAGVNAVKQTFPAKDLISAIAPKAEYOIKNTGELE 65  
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74  
QY 66 SLEMTKLEKDYDYLHMEYAVSLNDVSTFPEDESIDPLASLKOKIWKIPSGELIN 125  
DB 75 SIYEIMOKCALDYDELAKKEYTEKGLVYISTFSPAGANRLEDMGVSAFKISGSCNN 134  
QY 126 LPYIEKIAKLPIDPKKIISTGMAITIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185  
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188  
QY 186 DVNINAINDLKKHPKKNIGFSDHSGFYAAIAAVPYGITFEKHFTLDKSMGPDHLAS 245  
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTDNLACGVALGACVLEKHFITDSMERSGPDIVCS 247  
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVFSKNI 304  
DB 248 MDTOALKEILLIOSEQMAIMRGNNESKKAQEQVTIDPAPASVVISIKKGEVLSMDNI 307  
QY 305 TTKRPG-NGISPMWYNLLGKIAEQDFIPDELIHSEF 341  
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTQLSYEDF 345  
RESULT 11  
US-10-303-162-35  
Sequence 35, Application US/10303162  
Publication No. US20030157658A1  
GENERAL INFORMATION:  
APPLICANT: Gilbert, Michel  
APPLICANT: Wakarchuk, Warren W.  
TITLE OF INVENTION: Gangliobacter Glycosyltransferases for Biosynthesis of  
TITLE OF INVENTION: Gangliobacter Glycosyltransferases for Biosynthesis of  
FILE REFERENCE: 019633-000111US  
CURRENT FILING DATE: 2002-11-21  
PRIOR FILING DATE: 2001-03-21  
PRIOR FILING DATE: 2001-03-21  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 60/118,213  
PRIOR APPLICATION NUMBER: US 09/495,406  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 35  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a  
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
US-10-303-162-35  
Query Match 26.7%; Score 475.5; DB 14; Length 346;  
Best Local Similarity 36.7%; Pred. No. 1,1e-36;  
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;  
QY 6 IVAIGCHNNSVDIAREMILKAEAGVNAVKQTFPAKDLISAIAPKAEYOIKNTGELE 65  
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74  
QY 66 SLEMTKLEKDYDYLHMEYAVSLNDVSTFPEDESIDPLASLKOKIWKIPSGELIN 125  
DB 75 SIYEIMOKCALDYDELAKKEYTEKGLVYISTFSPAGANRLEDMGVSAFKISGSCNN 134  
QY 126 LPYIEKIAKLPIDPKKIISTGMAITIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185  
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188  
QY 186 DVNINAINDLKKHPKKNIGFSDHSGFYAAIAAVPYGITFEKHFTLDKSMGPDHLAS 245  
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTDNLACGVALGACVLEKHFITDSMERSGPDIVCS 247  
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVFSKNI 304  
DB 248 MDTOALKEILLIOSEQMAIMRGNNESKKAQEQVTIDPAPASVVISIKKGEVLSMDNI 307  
QY 305 TTKRPG-NGISPMWYNLLGKIAEQDFIPDELIHSEF 341  
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTQLSYEDF 345

QY 186 DVNINAINDLKKHPKKNIGFSDHSGFYAAIAAVPYGITFEKHFTLDKSMGPDHLAS 245  
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTDNLACGVALGACVLEKHFITDSMERSGPDIVCS 247  
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVFSKNI 304  
DB 248 MDTOALKEILLIOSEQMAIMRGNNESKKAQEQVTIDPAPASVVISIKKGEVLSMDNI 307  
QY 305 TTKRPG-NGISPMWYNLLGKIAEQDFIPDELIHSEF 341  
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTQLSYEDF 345  
RESULT 11  
US-10-303-162-35  
Sequence 35, Application US/10303162  
Publication No. US20030157658A1  
GENERAL INFORMATION:  
APPLICANT: Gilbert, Michel  
APPLICANT: Wakarchuk, Warren W.  
TITLE OF INVENTION: Gangliobacter Glycosyltransferases for Biosynthesis of  
TITLE OF INVENTION: Gangliobacter Glycosyltransferases for Biosynthesis of  
FILE REFERENCE: 019633-000111US  
CURRENT FILING DATE: 2002-11-21  
PRIOR FILING DATE: 2001-03-21  
PRIOR FILING DATE: 2001-03-21  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: US 60/118,213  
PRIOR APPLICATION NUMBER: US 09/495,406  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 35  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Campylobacter jejuni  
FEATURE:  
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a  
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
US-10-303-162-35  
Query Match 26.7%; Score 475.5; DB 14; Length 346;  
Best Local Similarity 36.7%; Pred. No. 1,1e-36;  
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;  
QY 6 IVAIGCHNNSVDIAREMILKAEAGVNAVKQTFPAKDLISAIAPKAEYOIKNTGELE 65  
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74  
QY 66 SLEMTKLEKDYDYLHMEYAVSLNDVSTFPEDESIDPLASLKOKIWKIPSGELIN 125  
DB 75 SIYEIMOKCALDYDELAKKEYTEKGLVYISTFSPAGANRLEDMGVSAFKISGSCNN 134  
QY 126 LPYIEKIAKLPIDPKKIISTGMAITIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185  
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188  
QY 186 DVNINAINDLKKHPKKNIGFSDHSGFYAAIAAVPYGITFEKHFTLDKSMGPDHLAS 245  
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTDNLACGVALGACVLEKHFITDSMERSGPDIVCS 247  
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVFSKNI 304  
DB 248 MDTOALKEILLIOSEQMAIMRGNNESKKAQEQVTIDPAPASVVISIKKGEVLSMDNI 307  
QY 305 TTKRPG-NGISPMWYNLLGKIAEQDFIPDELIHSEF 341  
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTQLSYEDF 345

Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 6

```
RESULT 12
US-10-156-761-10910
; Sequence 10910, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBAH, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10910
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10910

Query Match          24.0%; Score 426; DB 14; Length 311;
Best Local Similarity 33.9%; Pred. No. 5,1e-22;
Matches 100; Conservative 60; Mismatches 117; Indels 18; Gaps 6;

QY 4 IVVAEIGCHNGSVDIAREMILKAEAGVNAVYFQTFKADKLISALAPAEYQIKNT--61
DB 20 VYVVGELGINHNGELENAFLIDAAAEAGCDVAKFQ---KRTPEICTPPDDMDIERDTP 75
QY 62 -GELESQLEMTKLEMYVDYULHMEYAVSLNLDVFSPPDESDISIDFLASKQIMKIPS 120
DB 76 WGRW-TIIDYRHRREFEEDBYRQIDETAKSKNIDWPSPDITAAVAFLEKEDIPAKRVAS 134
QY 121 GELNLPYLERKIALPDPDKKIIISTGMATIDELIKOSVSIFNNKVPVGNITILHCNTEY 180
DB 135 ASLTDD--DELLRALRGTCRTVILISTGKSTPKQIRHAEVETASD-----NILCHATSTY 186
QY 181 PTPEDVNAINDLKHPKNNIGFSDHSGGYAAIAVPYGITFEKEFTLDKSMGDP 240
DB 187 PAKAEELNLKVINTLQAEFVNVPLIGSGHETGLQTLIAVALGANTFPERHITLIDRAMGS 246
QY 241 DHAASIEPDELKHLCTGVRCEKSLGNSHVTASE--RNKKIVARKSIIIAKTEI 293
DB 247 DQAAVPEPQGLTRLVDRDITETESLIGDGVKVESELGPMKKLRVGVAAEARI 301

RESULT 13
US-09-882-227-490
; Sequence 490, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Ooomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396a1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 490
```

```
LENGTH: 340
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-490

Query Match          22.6%; Score 402; DB 10; Length 340;
Best Local Similarity 31.8%; Pred. No. 1,1e-29;
Matches 104; Conservative 74; Mismatches 135; Indels 14; Gaps 8;

QY 6 IVAEIGCHNGSVDIAREMILKAEAGVNAVYFQTFKADKLISALAPAEYQIKNT--GEL 64
DB 7 IVVAEIGCHNGSVDIAREMILKAEAGVNAVYFQTFKADKLISALAPAEYQIKNT--GEL 65
QY 65 ESQLEMTKLEMYVDYULHMEYAVSLNLDVFSPPDESDISIDFLASKQIMKIPSGLT 124
DB 66 EULYELYOKASTPLEMHAELFELARKDLGIFSSPSSQALELLESLNCPWKIASFEIV 125
QY 125 NLPEYKXIAKLPIPPDKKIIISTGMATIDELIKOSVSIF--INNKPVGNITILHCNTEYPT 182
DB 126 DIDLIEKAR--TQKPIILSSGITHRELQDAISLCRRVNN---PITILKCVSAYPS 178
QY 183 PTPEDVNAINDLKHPKNNIGFSDHSGGYAAIAVPYGITFEKEFTLDKSMGDPH 242
DB 179 KIEDANLISMVYLGEIFGV--KGLSDHTIGSLCPILATTLGASMEKEKFIILKSKQTPPS 237
QY 243 LASIEPDELKHLCTGVRCEKSLGNS--KVVTASERKKIVARKSIIIAKTEIKGSPSE 301
DB 238 AFSWDFNGKSVKVAIKOSVTLAGEEPRINPKTIEKRRFARSLFVIR-DIQGEALTE 296
QY 302 KNITTRPGNGISPMWYNLGLKIAEQ 328
DB 297 NIKKALRPLGLHPKFEKELIGQKASK 323

RESULT 14
US-10-335-977-7626
; Sequence 7626, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBLACTER PYLOI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)742-4214
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7626:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
```

Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 7

TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...340  
SEQUENCE DESCRIPTION: SEQ ID NO: 7626  
US-10-335-977-7626

Query Match 21.4%; Score 381; DB 12; Length 340;  
Best Local Similarity 30.1%; Pred. No. 1,2e-27;  
Matches 98; Conservative 75; Mismatches 141; Indels 12; Gaps 6;

QY 6 IYAEIGCNHNSVDIAREMILKAKAGVNAVVFOTFKADKLISALAPKAEYQIKNT-GEL 64  
DB 7 IYAEISANHQDINLAKESLHAIKESGADPVKLQY-TPSQMTLDSKEDPFIIOGTLMDK 65  
QY 65 ESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSDIFLASLKQIKIPSGELL 124  
DB 66 ENLYGKQKASTPLEMHAELFELAKKLDLGISSPSSKALELSDCEMTKIASFEIV 125  
QY 125 NLEYLEKIKALPIPDKKIISTGMATIDEIKOSVSIF--INNKPVGNTIILHNTTEPT 182  
DB 126 DDLLEKAAAR--TQKPIILSSGIAHAELODAISLCRGVNN---FDITLLKCVSAFPS 178  
QY 183 PREDVNLAINDLKGFPPKNNIGFSQHSQFYAALAVPYGTFTEKHFTLDSKSGPDH 242  
DB 179 KIDDAHLASVVKLGTEFGV-KFGLSDHTIGSLCPILATTLGASVTEKHFTLNSLOTPDS 237  
QY 243 LASIPEDELKHLCTGVRCVCKSLGNSKVTASERKNIIVARSKIIAKTEIRKGEVSEK 302  
DB 238 AFSMDPNGFSKMWGALKOSVIALGEEBPKINPKTLERRPPARSLVFIYDIOGGEALTSD 287  
QY 303 NITTKRPGNGISPMETNLLGKIAEQ 328  
DB 298 NIKALRPNLGTHPKFYKEIILQOKASK 323

RESULT 15  
US-10-335-977-7625  
Sequence 7625, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEOTIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESS: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-Dec-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoudas, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7625:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 314 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...314  
SEQUENCE DESCRIPTION: SEQ ID NO: 7625:  
US-10-335-977-7625

Query Match 21.0%; Score 373; DB 12; Length 314;  
Best Local Similarity 30.4%; Pred. No. 6e-27;  
Matches 95; Conservative 72; Mismatches 134; Indels 12; Gaps 6;

QY 6 IYAEIGCNHNSVDIAREMILKAKAGVNAVVFOTFKADKLISALAPKAEYQIKNT-GEL 64  
DB 11 IYAEISANHQDINLAKESLHAIKESGADPVKLQY-TPSQMTLDSKEDPFIIOGTLMDK 69  
QY 65 ESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSDIFLASLKQIKIPSGELL 124  
DB 70 ENLYGKQKASTPLEMHAELFELAKKLDLGISSPSSKALELSDCEMTKIASFEIV 129  
QY 125 NLEYLEKIKALPIPDKKIISTGMATIDEIKOSVSIF--INNKPVGNTIILHNTTEPT 182  
DB 130 DDLLEKAAAR--TQKPIILSSGIAHAELODAISLCRGVNN---FDITLLKCVSAFPS 182  
QY 183 PREDVNLAINDLKGFPPKNNIGFSQHSQFYAALAVPYGTFTEKHFTLDSKSGPDH 242  
DB 183 KIDDAHLASVVKLGTEFGV-KFGLSDHTIGSLCPILATTLGASVTEKHFTLNSLOTPDS 241  
QY 243 LASIPEDELKHLCTGVRCVCKSLGNSKVTASERKNIIVARSKIIAKTEIRKGEVSEK 302  
DB 242 AFSMDPNGFSKMWGALKOSVIALGEEBPKINPKTLERRPPARSLVFIYDIOGGEALTSD 301  
QY 303 NITTKRPGNGISP 315  
DB 302 NIKALRPNLGTVT 314

Search completed: May 6, 2004, 09:25:08  
Job time: 38.4054 secs